524 Rec'd PCT/PTO 26 OCT 1999

FORM-PTO-1390 (Rev. 10-96) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER: 030708-035

# TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

U.S. APPLICATION NO. (If known, see 37 C.F.R. 1.5): 09/403724

| INTERNATIONAL APPLICATION NO. | INTERNATIONAL FILING DATE | PRIORITY DATE CLAIMED |
|---|---|---|
| PCT/IB98/00625 | 24 April 1998 | 26 April 1997 |

TITLE OF INVENTION: NEUROTRYPSIN

APPLICANT(S) FOR DO/EO/US: Peter Sonderegger

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. [X] This is an express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and the PCT Articles 22 and 39(1).
4. [ ] A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. [X] is transmitted herewith (required only if not transmitted by the International Bureau).
   b. [X] has been transmitted by the International Bureau.
   c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US)
6. [ ] A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. [ ] Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. [ ] are transmitted herewith (required only if not transmitted by the International Bureau).
   b. [ ] have been transmitted by the International Bureau.
   c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
   d. [ ] have not been made and will not be made.
8. [ ] A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. [ ] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

**Items 11. to 16. below concern other document(s) or information included:**

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
    [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [X] A FIRST preliminary amendment.
    [ ] A SECOND or SUBSEQUENT preliminary amendment.
14. [ ] A substitute specification.
15. [ ] A change of power of attorney and/or address letter.
16. [ ] Other items or information:

420 Rec'd PCT/PTO 26 OCT 1999

| U.S. APPLICATION NO. (If known, see 37 CFR 1.5) | INTERNATIONAL APPLICATION NO. | ATTORNEY'S DOCKET NUMBER |
|---|---|---|
| 09/403724 | PCT/IB98/00625 | 030708-035 |

| 17. ☒ The following fees are submitted: | CALCULATIONS | PTO USE ONLY |
|---|---|---|
| **Basic National Fee (37 CFR 1.492(a)(1)-(5)):** | | |
| Search Report has been prepared by the EPO or JPO ..... $840.00 (970) | | |
| International preliminary examination fee paid to USPTO (37 CFR 1.482) ..... $670.00 (956) | | |
| No international preliminary examination fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2)) ..... $760.00 (958) | | |
| Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO ..... $970.00 (960) | | |
| International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4) ..... $96.00 (962) | | |
| **ENTER APPROPRIATE BASIC FEE AMOUNT =** | $ 970.00 | |
| Surcharge of **$130.00 (154)** for furnishing the oath or declaration later than 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)). | $ 0.00 | |

| Claims | Number Filed | Number Extra | Rate | | |
|---|---|---|---|---|---|
| Total Claims | 15 -20 = | 0 | X$18.00 (966) | $ 0.00 | |
| Independent Claims | 14 -3 = | 11 | X$78.00 (964) | $ 858.00 | |
| Multiple dependent claim(s) (if applicable) | | | +$260.00 (968) | $ 0.00 | |
| **TOTAL OF ABOVE CALCULATIONS =** | | | | $ 1,828.00 | |
| Reduction for 1/2 for filing by small entity, if applicable. Verified Small Entity statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28). | | | | $ - | |
| **SUBTOTAL =** | | | | $ | |
| Processing fee of **$130.00 (156)** for furnishing the English translation later than 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)). + | | | | $ | |
| **TOTAL NATIONAL FEE =** | | | | $ 1,828.00 | |
| Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). per property + | | | | $ | |
| **TOTAL FEES ENCLOSED =** | | | | $ 1,828.00 | |
| | | | | Amount to be: refunded | $ |
| | | | | charged | $ |

a. [X] A check in the amount of $ 1,828.00 to cover the above fees is enclosed.

b. [ ] Please charge my Deposit Account No. 02-4800 in the amount of $_____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. [X] The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 02-4800. A duplicate copy of this sheet is enclosed.

**NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

William L. Mathis
BURNS, DOANE, SWECKER & MATHIS, L.L.P.
P.O. Box 1404
Alexandria, Virginia 22313-1404

Bruce J. Boggs, Jr.
SIGNATURE
Reg No. 37,027
Bruce J. Boggs, Jr.
NAME

32,344
REGISTRATION NUMBER

Patent
Attorney's Docket No. 030708-035

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Patent Application of )
)
Peter SONDEREGGER )
)
Serial No.: 09/403,724 )
)
Filed: October 26, 1999 )
)
For: NEUROTRYPSIN )


Group Art Unit: Unknown

Examiner: Unknown

**ATTENTION: BOX SEQUENCE**

**TRANSMITTAL LETTER FOR MISSING PARTS OF APPLICATION**

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

In complete response to the Notice to Comply with Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Sequence disclosures dated **not yet received**, enclosed please find:

[X] A copy of the "Sequence Listing" in computer readable form in compliance with 37 C.F.R. §§1.823(b) and 1.824.

[X] A statement that the content of the paper and computer readable copies are the same as set forth in 37 C.F.R. §1.821(f).

The Commissioner is hereby authorized to charge any additional fees under 37 C.F.R. §§1.16, 1.17, and 1.21 that may be required by this paper, and to credit any overpayment to Deposit Account No. 02-4800. A duplicate copy of this paper is enclosed.

Respectfully submitted,

1737 King Street, Suite 500
Alexandria, VA 22314-2756
(703) 836-6620

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

Date: December 20, 1999

By _____
Richard C. Ekstrom
Registration No. 37,027

#4

Patent
Attorney's Docket No. 030708-035

Applicant or Patentee: Peter Sonderegger

Application or Patent No.:

Filed or Issued: October 26, 1999

For: NEUROTRYPSIN


## VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS (37 C.F.R. §§ 1.9(f) AND 1.27(b)) - INDEPENDENT INVENTOR

As a below-named inventor, I hereby declare that I qualify as an independent inventor as defined in 37 C.F.R. § 1.9(c) for purposes of paying reduced fees under Sections 41(a) and 41(b) of Title 35, United States Code, to the Patent and Trademark Office with regard to the invention entitled Neurotrypsin described in:

[ ] the specification filed herewith
[X] Application No. _____, filed October 26, 1999.
[ ] Patent No. _____, issued _____.

I have not assigned, granted, conveyed, or licensed and am under no obligation under contract or law to assign, grant, convey, or license any rights in the invention either to any person who could not be classified as an independent inventor under 37 C.F.R. § 1.9(c) if that person had made the invention, or to any concern that would not qualify as either a small business concern under 37 C.F.R. § 1.9(d) or a nonprofit organization under 37 C.F.R. § 1.9(e).

Each person, concern or organization to which I have assigned, granted, conveyed, or licensed or am under an obligation under contract or law to assign, grant, convey, or license any rights in the invention is listed below:

[X] no such person, concern, or organization
[ ] persons, concerns, or organizations listed below*

*NOTE: Separate verified statements are required from each named person, concern, or organization having rights to the invention averring to their status as small entities. (37 C.F.R. § 1.27.)

FULL NAME _____

ADDRESS _____

[ ] individual [ ] small business concern [ ] nonprofit organization

FULL NAME _____

ADDRESS _____

[ ] individual [ ] small business concern [ ] nonprofit organization

FULL NAME _____

ADDRESS _____

[ ] individual [ ] small business concern [ ] nonprofit organization

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earlier of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 C.F.R. § 1.28(b).)

Application No. _____
Attorney's Docket No. 030708-035

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code; and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

Name Peter Sonderegger

Signature P. Sonderegger Date Nov-11-1999

Patent
Attorney's Docket No. 030708-035

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

| | | |
|---|---|---|
| In re Patent Application of | | |
| Peter SONDEREGGER | | Group Art Unit: Unassigned |
| Serial No.: 09/403,724 | | Examiner: Unassigned |
| Filed: October 26, 1999 | | **ATTENTION: BOX SEQUENCE** |
| For: NEUROTRYPSIN | | |



**PRELIMINARY AMENDMENT**

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Prior to examination on the merits, please amend the above-identified application as follows:

IN THE SPECIFICATION:

In compliance with 37 C.F.R. §1.823(a), please delete pages 16-32 of the specification and insert therefor the attached paper copy of the "Sequence Listing" between page 15 of the Disclosure and the first page of the Claims to replace the Sequence Listing identified thereon.

REMARKS

The paper copy of the Sequence Listing for the subject application, is by this amendment added between page 15 of the Specification and the first page of the Claims to replace the Sequence Listing identified thereon. Please amend the page numbers accordingly.

Favorable consideration on the merits is respectfully requested.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

By _____
Richard C. Ekstrom
Registration No. 37,027

P.O. Box 1404
Alexandria, Virginia 22313-1404
(703) 836-6620

Date: December 20, 1999

SEQUENCE LISTING

<110> SONDEREGGER, Peter

<120> NEUROTRYPSIN

<130> 030708-035

<140> 09/403,724
<141> 1999-10-26

<150> PCT/IB98/00625
<151> 1998-04-24

<150> CH 0966/97
<151> 1997-04-26

<160> 28

<170> PatentIn Ver. 2.0

<210> 1
<211> 3350
<212> DNA
<213> Homo sapiens

<220>
<221> sig_peptide
<222> (44)..(103)

<220>
<221> mat_peptide
<222> (104)..(2668)

<220>
<221> CDS
<222> (44)..(2668)

<220>
<221> 5'UTR
<222> (1)..(43)

<220>
<221> 3'UTR
<222> (2669)..(3350)

<400> 1
cggaagctgg ggagcatgga ccagaccccg cagcgctggc acc atg acg ctc gcc 55
Met Thr Leu Ala
-20

cgc ttc gtg cta gcc ctg atg tta ggg gcg ctc ccc gaa gtg gtc ggc 103
Arg Phe Val Leu Ala Leu Met Leu Gly Ala Leu Pro Glu Val Val Gly
-15 -10 -5 -1

```
ttt gat tct gtc ctc aat gat tcc ctc cac cac agc cac cgc cat tcg   151
Phe Asp Ser Val Leu Asn Asp Ser Leu His His Ser His Arg His Ser
  1               5                  10                  15

ccc cct gcg ggt ccg cac tac ccc tat tac ctt ccc acc cag cag cgg   199
Pro Pro Ala Gly Pro His Tyr Pro Tyr Tyr Leu Pro Thr Gln Gln Arg
            20                  25                  30

ccc ccg acg acg cgt ccg ccg ccg cct ctc ccg cgc ttc ccg cgc ccc   247
Pro Pro Thr Thr Arg Pro Pro Pro Pro Leu Pro Arg Phe Pro Arg Pro
        35                  40                  45

ccg cgg gcg ctc cct gcc cag cgc ccg cac gcc ctc cag gcc ggg cac   295
Pro Arg Ala Leu Pro Ala Gln Arg Pro His Ala Leu Gln Ala Gly His
    50                  55                  60

acg ccc cgg ccg cac ccc tgg ggc tgc ccc gcc ggc gag cca tgg gtc   343
Thr Pro Arg Pro His Pro Trp Gly Cys Pro Ala Gly Glu Pro Trp Val
65                  70                  75                  80

agc gtg acg gac ttc ggc gcc ccg tgt ctg cgg tgg gcg gag gtg cca   391
Ser Val Thr Asp Phe Gly Ala Pro Cys Leu Arg Trp Ala Glu Val Pro
                85                  90                  95

ccc ttc ctg gag cgg tcg ccc cca gcg agc tgg gct cag ctg cga gga   439
Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Gln Leu Arg Gly
            100                 105                 110

cag cgc cac aac ttt tgt cgg agc ccc gac ggc gcg ggc aga ccc tgg   487
Gln Arg His Asn Phe Cys Arg Ser Pro Asp Gly Ala Gly Arg Pro Trp
        115                 120                 125

tgt ttc tac gga gac gcc cgt ggc aag gtg gac tgg ggc tac tgc gac   535
Cys Phe Tyr Gly Asp Ala Arg Gly Lys Val Asp Trp Gly Tyr Cys Asp
    130                 135                 140

tgc aga cac gga tca gta cga ctt cgt ggc ggc aaa aat gag ttt gaa   583
Cys Arg His Gly Ser Val Arg Leu Arg Gly Gly Lys Asn Glu Phe Glu
145                 150                 155                 160

ggc aca gtg gaa gta tat gca agt gga gtt tgg ggc act gtc tgt agc   631
Gly Thr Val Glu Val Tyr Ala Ser Gly Val Trp Gly Thr Val Cys Ser
                165                 170                 175

agc cac tgg gat gat tct gat gca tca gtc att tgt cac cag ctg cag   679
Ser His Trp Asp Asp Ser Asp Ala Ser Val Ile Cys His Gln Leu Gln
            180                 185                 190

ctg gga gga aaa gga ata gca aaa caa acc ccg ttt tct gga ctg ggc   727
Leu Gly Gly Lys Gly Ile Ala Lys Gln Thr Pro Phe Ser Gly Leu Gly
        195                 200                 205

ctt att ccc att tat tgg agc aat gtc cgt tgc cga gga gat gaa gaa   775
Leu Ile Pro Ile Tyr Trp Ser Asn Val Arg Cys Arg Gly Asp Glu Glu
    210                 215                 220
```

```
aat ata ctg ctt tgt gaa aaa gac atc tgg cag ggt ggg gtg tgt cct   823
Asn Ile Leu Leu Cys Glu Lys Asp Ile Trp Gln Gly Gly Val Cys Pro
225                 230                 235                 240

cag aag atg gca gct gct gtc acg tgt agc ttt tcc cat ggc cca acg   871
Gln Lys Met Ala Ala Ala Val Thr Cys Ser Phe Ser His Gly Pro Thr
                245                 250                 255

ttc ccc atc att cgc ctt gct gga ggc agc agt gtg cat gaa ggc cgg   919
Phe Pro Ile Ile Arg Leu Ala Gly Gly Ser Ser Val His Glu Gly Arg
            260                 265                 270

gtg gag ctc tac cat gct ggc cag tgg gga acc gtt tgt gat gac caa   967
Val Glu Leu Tyr His Ala Gly Gln Trp Gly Thr Val Cys Asp Asp Gln
        275                 280                 285

tgg gat gat gcc gat gca gaa gtg atc tgc agg cag ctg ggc ctc agt  1015
Trp Asp Asp Ala Asp Ala Glu Val Ile Cys Arg Gln Leu Gly Leu Ser
    290                 295                 300

ggc att gcc aaa gca tgg cat cag gca tat ttt ggg gaa ggg tct ggc  1063
Gly Ile Ala Lys Ala Trp His Gln Ala Tyr Phe Gly Glu Gly Ser Gly
305                 310                 315                 320

cca gtt atg ttg gat gaa gta cgc tgc act ggg aat gag ctt tca att  1111
Pro Val Met Leu Asp Glu Val Arg Cys Thr Gly Asn Glu Leu Ser Ile
                325                 330                 335

gag cag tgt cca aag agc tcc tgg gga gag cat aac tgt ggc cat aaa  1159
Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn Cys Gly His Lys
            340                 345                 350

gaa gat gct gga gtg tcc tgt acc cct cta aca gat ggg gtc atc aga  1207
Glu Asp Ala Gly Val Ser Cys Thr Pro Leu Thr Asp Gly Val Ile Arg
        355                 360                 365

ctt gca ggt ggg aaa ggc agc cat gag ggt cgc ttg gag gta tat tac  1255
Leu Ala Gly Gly Lys Gly Ser His Glu Gly Arg Leu Glu Val Tyr Tyr
    370                 375                 380

aga ggc cag tgg gga act gtc tgt gat gat ggc tgg act gag ctg aat  1303
Arg Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp Thr Glu Leu Asn
385                 390                 395                 400

aca tac gtg gtt tgt cga cag ttg gga ttt aaa tat ggt aaa caa gca  1351
Thr Tyr Val Val Cys Arg Gln Leu Gly Phe Lys Tyr Gly Lys Gln Ala
                405                 410                 415

tct gcc aac cat ttt gaa gaa agc aca ggg ccc ata tgg ttg gat gac  1399
Ser Ala Asn His Phe Glu Glu Ser Thr Gly Pro Ile Trp Leu Asp Asp
            420                 425                 430

gtc agc tgc tca gga aag gaa acc aga ttt ctt cag tgt tcc agg cga  1447
Val Ser Cys Ser Gly Lys Glu Thr Arg Phe Leu Gln Cys Ser Arg Arg
        435                 440                 445
```

```
cag tgg gga agg cat gac tgc agc cac cgc gaa gat gtt agc att gcc   1495
Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp Val Ser Ile Ala
    450                 455                 460

tgc tac cct ggc ggc gag gga cac agg ctc tct ctg ggt ttt cct gtc   1543
Cys Tyr Pro Gly Gly Glu Gly His Arg Leu Ser Leu Gly Phe Pro Val
465                 470                 475                 480

aga ctg atg gat gga gaa aat aag aaa gaa gga cga gtg gag gtt ttt   1591
Arg Leu Met Asp Gly Glu Asn Lys Lys Glu Gly Arg Val Glu Val Phe
                485                 490                 495

atc aat ggc cag tgg gga aca atc tgt gat gat gga tgg act gat aag   1639
Ile Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly Trp Thr Asp Lys
            500                 505                 510

gat gca gct gtg atc tgt cgt cag ctt ggc tac aag ggt cct gcc aga   1687
Asp Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys Gly Pro Ala Arg
        515                 520                 525

gca aga acc atg gct tac ttt gga gaa gga aaa gga ccc atc cat gtg   1735
Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly Pro Ile His Val
    530                 535                 540

gat aat gtg aag tgc aca gga aat gag agg tcc ttg gct gac tgt atc   1783
Asp Asn Val Lys Cys Thr Gly Asn Glu Arg Ser Leu Ala Asp Cys Ile
545                 550                 555                 560

aag caa gat att gga aga cac aac tgc cgc cac agt gaa gat gca gga   1831
Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser Glu Asp Ala Gly
                565                 570                 575

gtt att tgt gat tat ttt ggc aag aag gcc tca ggt aac agt aat aaa   1879
Val Ile Cys Asp Tyr Phe Gly Lys Lys Ala Ser Gly Asn Ser Asn Lys
            580                 585                 590

gag tcc ctc tca tct gtt tgt ggc ttg aga tta ctg cac cgt cgg cag   1927
Glu Ser Leu Ser Ser Val Cys Gly Leu Arg Leu Leu His Arg Arg Gln
        595                 600                 605

aag cgg atc att ggt ggg aaa aat tct tta agg ggt ggt tgg cct tgg   1975
Lys Arg Ile Ile Gly Gly Lys Asn Ser Leu Arg Gly Gly Trp Pro Trp
    610                 615                 620

cag gtt tcc ctc cgg ctg aag tca tcc cat gga gat ggc agg ctc ctc   2023
Gln Val Ser Leu Arg Leu Lys Ser Ser His Gly Asp Gly Arg Leu Leu
625                 630                 635                 640

tgc ggg gct acg ctc ctg agt agc tgc tgg gtc ctc aca gca gca cac   2071
Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu Thr Ala Ala His
                645                 650                 655

tgt ttc aag agg tat ggc aac agc act agg agc tat gct gtt agg gtt   2119
Cys Phe Lys Arg Tyr Gly Asn Ser Thr Arg Ser Tyr Ala Val Arg Val
            660                 665                 670
```

```
gga gat tat cat act ctg gta cca gag gag ttt gag gaa gaa att gga   2167
Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu Glu Glu Ile Gly
        675                 680                 685

gtt caa cag att gtg att cat cgg gag tat cga ccc gac cgc agt gat   2215
Val Gln Gln Ile Val Ile His Arg Glu Tyr Arg Pro Asp Arg Ser Asp
    690                 695                 700

tat gac ata gcc ctg gtt aga tta caa gga cca gaa gag caa tgt gcc   2263
Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Glu Glu Gln Cys Ala
705                 710                 715                 720

aga ttc agc agc cat gtt ttg cca gcc tgt tta cca ctc tgg aga gag   2311
Arg Phe Ser Ser His Val Leu Pro Ala Cys Leu Pro Leu Trp Arg Glu
                725                 730                 735

agg cca cag aaa aca gca tcc aac tgt tac ata aca gga tgg ggt gac   2359
Arg Pro Gln Lys Thr Ala Ser Asn Cys Tyr Ile Thr Gly Trp Gly Asp
            740                 745                 750

aca gga cga gcc tat tca aga aca cta caa caa gca gcc att ccc tta   2407
Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala Ala Ile Pro Leu
        755                 760                 765

ctt cct aaa agg ttt tgt gaa gaa cgt tat aag ggt cgg ttt aca ggg   2455
Leu Pro Lys Arg Phe Cys Glu Glu Arg Tyr Lys Gly Arg Phe Thr Gly
    770                 775                 780

aga atg ctt tgt gct gga aac ctc cat gaa cac aaa cgc gtg gac agc   2503
Arg Met Leu Cys Ala Gly Asn Leu His Glu His Lys Arg Val Asp Ser
785                 790                 795                 800

tgc cag gga gac agc gga gga cca ctc atg tgt gaa cgg ccc gga gag   2551
Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu Arg Pro Gly Glu
                805                 810                 815

agc tgg gtg gtg tat ggg gtg acc tcc tgg ggg tat ggc tgt gga gtc   2599
Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr Gly Cys Gly Val
            820                 825                 830

aag gat tct cct ggt gtt tat acc aaa gtc tca gcc ttt gta cct tgg   2647
Lys Asp Ser Pro Gly Val Tyr Thr Lys Val Ser Ala Phe Val Pro Trp
        835                 840                 845

ata aaa agt gtc acc aaa ctg taattcttca tggaaacttc aaagcagcat      2698
Ile Lys Ser Val Thr Lys Leu
    850                 855

ttaaacaaat ggaaaacttt gaacccccac tattagcact cagcagagat gacaacaaat 2758

ggcaagatct gtttttgctt tgtgttgtgg taaaaaattg tgtacccccc gctgcttttg 2818

agaaatttgt gaacattttc agaggcctca gtgtagtgga agtgataatc cttaaatgaa 2878

cattttctac cctaatttca ctggagtgac ttattctaag cctcatctat cccctaccta 2938
```

tttctcaaaa tcattctatg ctgattttac aaaagatcat ttttacattt gaactgagaa 2998

ccccttttaa ttgaatcagt ggtgtctgaa atcatattaa atacccacat ttgacataaa 3058

tgcggtaccc tttactacac tcatgagtgg catatttatg cttaggtctt ttcaaaagac 3118

ttgacaagaa atcttcatat tctctgtagc ctttgtcaag tgaggaaatc agtggttaaa 3178

gaattccact ataaactttt aggcctgaat aggagtagta aagcctcaag gacatctgcc 3238

tgtcacaata tattctcaaa gtgatctgat atttggaaac aagtatcctt gttgagtacc 3298

aagtgctaca gaaaccataa gataaaaata ctttctacct acagcgtgcc cg 3350

<210> 2
<211> 875
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Thr Leu Ala Arg Phe Val Leu Ala Leu Met Leu Gly Ala Leu Pro
-20                 -15                 -10                  -5

Glu Val Val Gly Phe Asp Ser Val Leu Asn Asp Ser Leu His His Ser
            -1   1                5                  10

His Arg His Ser Pro Pro Ala Gly Pro His Tyr Pro Tyr Tyr Leu Pro
        15                  20                  25

Thr Gln Gln Arg Pro Pro Thr Thr Arg Pro Pro Pro Pro Leu Pro Arg
    30                  35                  40

Phe Pro Arg Pro Pro Arg Ala Leu Pro Ala Gln Arg Pro His Ala Leu
45                  50                  55                  60

Gln Ala Gly His Thr Pro Arg Pro His Pro Trp Gly Cys Pro Ala Gly
                65                  70                  75

Glu Pro Trp Val Ser Val Thr Asp Phe Gly Ala Pro Cys Leu Arg Trp
            80                  85                  90

Ala Glu Val Pro Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala
        95                  100                 105

Gln Leu Arg Gly Gln Arg His Asn Phe Cys Arg Ser Pro Asp Gly Ala
    110                 115                 120

Gly Arg Pro Trp Cys Phe Tyr Gly Asp Ala Arg Gly Lys Val Asp Trp
125                 130                 135                 140

Gly Tyr Cys Asp Cys Arg His Gly Ser Val Arg Leu Arg Gly Gly Lys
                145                 150                 155

Asn Glu Phe Glu Gly Thr Val Glu Val Tyr Ala Ser Gly Val Trp Gly
```

```
            160                 165                 170

Thr Val Cys Ser Ser His Trp Asp Asp Ser Asp Ala Ser Val Ile Cys
        175                 180                 185

His Gln Leu Gln Leu Gly Gly Lys Gly Ile Ala Lys Gln Thr Pro Phe
    190                 195                 200

Ser Gly Leu Gly Leu Ile Pro Ile Tyr Trp Ser Asn Val Arg Cys Arg
205                 210                 215                 220

Gly Asp Glu Glu Asn Ile Leu Leu Cys Glu Lys Asp Ile Trp Gln Gly
                225                 230                 235

Gly Val Cys Pro Gln Lys Met Ala Ala Ala Val Thr Cys Ser Phe Ser
            240                 245                 250

His Gly Pro Thr Phe Pro Ile Ile Arg Leu Ala Gly Gly Ser Ser Val
        255                 260                 265

His Glu Gly Arg Val Glu Leu Tyr His Ala Gly Gln Trp Gly Thr Val
    270                 275                 280

Cys Asp Asp Gln Trp Asp Asp Ala Asp Ala Glu Val Ile Cys Arg Gln
285                 290                 295                 300

Leu Gly Leu Ser Gly Ile Ala Lys Ala Trp His Gln Ala Tyr Phe Gly
                305                 310                 315

Glu Gly Ser Gly Pro Val Met Leu Asp Glu Val Arg Cys Thr Gly Asn
            320                 325                 330

Glu Leu Ser Ile Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn
        335                 340                 345

Cys Gly His Lys Glu Asp Ala Gly Val Ser Cys Thr Pro Leu Thr Asp
    350                 355                 360

Gly Val Ile Arg Leu Ala Gly Gly Lys Gly Ser His Glu Gly Arg Leu
365                 370                 375                 380

Glu Val Tyr Tyr Arg Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp
                385                 390                 395

Thr Glu Leu Asn Thr Tyr Val Val Cys Arg Gln Leu Gly Phe Lys Tyr
            400                 405                 410

Gly Lys Gln Ala Ser Ala Asn His Phe Glu Glu Ser Thr Gly Pro Ile
        415                 420                 425

Trp Leu Asp Asp Val Ser Cys Ser Gly Lys Glu Thr Arg Phe Leu Gln
    430                 435                 440

Cys Ser Arg Arg Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp
445                 450                 455                 460
```

```
Val Ser Ile Ala Cys Tyr Pro Gly Gly Glu Gly His Arg Leu Ser Leu
                465                 470                 475

Gly Phe Pro Val Arg Leu Met Asp Gly Glu Asn Lys Lys Glu Gly Arg
            480                 485                 490

Val Glu Val Phe Ile Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly
        495                 500                 505

Trp Thr Asp Lys Asp Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys
    510                 515                 520

Gly Pro Ala Arg Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly
525                 530                 535                 540

Pro Ile His Val Asp Asn Val Lys Cys Thr Gly Asn Glu Arg Ser Leu
                545                 550                 555

Ala Asp Cys Ile Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser
            560                 565                 570

Glu Asp Ala Gly Val Ile Cys Asp Tyr Phe Gly Lys Lys Ala Ser Gly
        575                 580                 585

Asn Ser Asn Lys Glu Ser Leu Ser Ser Val Cys Gly Leu Arg Leu Leu
    590                 595                 600

His Arg Arg Gln Lys Arg Ile Ile Gly Gly Lys Asn Ser Leu Arg Gly
605                 610                 615                 620

Gly Trp Pro Trp Gln Val Ser Leu Arg Leu Lys Ser Ser His Gly Asp
                625                 630                 635

Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu
            640                 645                 650

Thr Ala Ala His Cys Phe Lys Arg Tyr Gly Asn Ser Thr Arg Ser Tyr
        655                 660                 665

Ala Val Arg Val Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu
    670                 675                 680

Glu Glu Ile Gly Val Gln Gln Ile Val Ile His Arg Glu Tyr Arg Pro
685                 690                 695                 700

Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Glu
                705                 710                 715

Glu Gln Cys Ala Arg Phe Ser Ser His Val Leu Pro Ala Cys Leu Pro
            720                 725                 730

Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala Ser Asn Cys Tyr Ile Thr
        735                 740                 745

Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala
```

```
    750                 755                 760

Ala Ile Pro Leu Leu Pro Lys Arg Phe Cys Glu Glu Arg Tyr Lys Gly
765                 770                 775                 780

Arg Phe Thr Gly Arg Met Leu Cys Ala Gly Asn Leu His Glu His Lys
                785                 790                 795

Arg Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu
            800                 805                 810

Arg Pro Gly Glu Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr
        815                 820                 825

Gly Cys Gly Val Lys Asp Ser Pro Gly Val Tyr Thr Lys Val Ser Ala
    830                 835                 840

Phe Val Pro Trp Ile Lys Ser Val Thr Lys Leu
845                 850                 855
```

<210> 3
<211> 2356
<212> DNA
<213> Mus musculus

<220>
<221> sig_peptide
<222> (24)..(86)

<220>
<221> mat_peptide
<222> (87)..(2306)

<220>
<221> CDS
<222> (24)..(2306)

<220>
<221> polyA_site
<222> one-of(2341, 2356)

<220>
<221> 5'UTR
<222> (1)..(23)

<220>
<221> 3'UTR
<222> (2307)..one-of(2341, 2356)

<400> 3
ggaccacact cggcgccgca gcc atg gcg ctc gcc cgc tgc gtg ctg gct gtg 53
                        Met Ala Leu Ala Arg Cys Val Leu Ala Val
                            -20                 -15

```
att tta ggg gca ctg tct gta gtg gcc cgc gct gat ccg gtc tcg cgc   101
Ile Leu Gly Ala Leu Ser Val Val Ala Arg Ala Asp Pro Val Ser Arg
    -10                 -5                  -1   1                5

tct ccc ctt cac cgc ccg cat ccg tcc cca ccg cgt tcc caa cac gcg   149
Ser Pro Leu His Arg Pro His Pro Ser Pro Pro Arg Ser Gln His Ala
                10                  15                  20

cac tac ctt ccc agc tcg cgg cgg cca ccc agg acc ccg cgc ttc ccg   197
His Tyr Leu Pro Ser Ser Arg Arg Pro Pro Arg Thr Pro Arg Phe Pro
            25                  30                  35

ctc ccg ctg cgg atc ccc gct gcc cag cgc ccg cag gtc ctc agc acc   245
Leu Pro Leu Arg Ile Pro Ala Ala Gln Arg Pro Gln Val Leu Ser Thr
        40                  45                  50

ggg cac acg ccc ccg acg att cca cgc cgc tgc ggg gca gga gag tcg   293
Gly His Thr Pro Pro Thr Ile Pro Arg Arg Cys Gly Ala Gly Glu Ser
    55                  60                  65

tgg ggc aat gcc acc aac ctc ggc gtc ccg tgt cta cac tgg gac gag   341
Trp Gly Asn Ala Thr Asn Leu Gly Val Pro Cys Leu His Trp Asp Glu
70                  75                  80                  85

gtg ccg ccc ttc ctg gag cgg tcg ccc ccg gcc agt tgg gct gag ctg   389
Val Pro Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Glu Leu
                90                  95                  100

cga ggg cag ccg cac aac ttc tgc cgg agc ccg gat ggc tcg ggc aga   437
Arg Gly Gln Pro His Asn Phe Cys Arg Ser Pro Asp Gly Ser Gly Arg
            105                 110                 115

cct tgg tgc ttc tat cgg aat gcc cag ggc aaa gta gac tgg ggc tac   485
Pro Trp Cys Phe Tyr Arg Asn Ala Gln Gly Lys Val Asp Trp Gly Tyr
        120                 125                 130

tgc gat tgt ggt caa ggc ccg gcg ttg ccc gtc att cgc ctt gtt ggt   533
Cys Asp Cys Gly Gln Gly Pro Ala Leu Pro Val Ile Arg Leu Val Gly
    135                 140                 145

ggg aac agt ggg cat gaa ggt cga gtg gag ctg tac cac gct ggc cag   581
Gly Asn Ser Gly His Glu Gly Arg Val Glu Leu Tyr His Ala Gly Gln
150                 155                 160                 165

tgg ggg acc atc tgt gac gac caa tgg gac aat gca gac gca gac gtc   629
Trp Gly Thr Ile Cys Asp Asp Gln Trp Asp Asn Ala Asp Ala Asp Val
                170                 175                 180

atc tgt agg cag ctg ggg ctc agt ggc att gcc aaa gca tgg cat cag   677
Ile Cys Arg Gln Leu Gly Leu Ser Gly Ile Ala Lys Ala Trp His Gln
            185                 190                 195

gca cat ttt ggg gaa gga tct ggc cca ata ttg ttg gat gaa gta cgc   725
Ala His Phe Gly Glu Gly Ser Gly Pro Ile Leu Leu Asp Glu Val Arg
        200                 205                 210
```

```
tgc acc gga aac gag ctg tca att gag caa tgt cca aag agt tcc tgg   773
Cys Thr Gly Asn Glu Leu Ser Ile Glu Gln Cys Pro Lys Ser Ser Trp
    215                 220                 225

ggc gaa cat aac tgt ggc cat aaa gaa gat gct gga gtg tct tgt gtt   821
Gly Glu His Asn Cys Gly His Lys Glu Asp Ala Gly Val Ser Cys Val
230                 235                 240                 245

cct cta aca gat ggt gtc atc aga ctg gca gga gga aaa agt acc cat   869
Pro Leu Thr Asp Gly Val Ile Arg Leu Ala Gly Gly Lys Ser Thr His
                250                 255                 260

gaa ggt cgc ctg gag gtc tac tac aag ggg cag tgg ggg aca gtc tgt   917
Glu Gly Arg Leu Glu Val Tyr Tyr Lys Gly Gln Trp Gly Thr Val Cys
            265                 270                 275

gat gat ggc tgg act gag atg aac aca tac gtg gct tgt cga ctg ctg   965
Asp Asp Gly Trp Thr Glu Met Asn Thr Tyr Val Ala Cys Arg Leu Leu
        280                 285                 290

gga ttt aaa tac ggc aaa cag tcc tct gtg aac cat ttt gat ggc agc   1013
Gly Phe Lys Tyr Gly Lys Gln Ser Ser Val Asn His Phe Asp Gly Ser
    295                 300                 305

aac agg ccc ata tgg ctg gat gac gtc agc tgc tca gga aaa gaa gtc   1061
Asn Arg Pro Ile Trp Leu Asp Asp Val Ser Cys Ser Gly Lys Glu Val
310                 315                 320                 325

agc ttc att cag tgt tcc agg aga cag tgg gga agg cat gac tgc agc   1109
Ser Phe Ile Gln Cys Ser Arg Arg Gln Trp Gly Arg His Asp Cys Ser
                330                 335                 340

cat aga gaa gat gtg ggc ctc acc tgc tat cct gac agc gat gga cat   1157
His Arg Glu Asp Val Gly Leu Thr Cys Tyr Pro Asp Ser Asp Gly His
            345                 350                 355

agg ctt tct cca ggt ttt ccc atc aga cta gtg gat gga gag aat aag   1205
Arg Leu Ser Pro Gly Phe Pro Ile Arg Leu Val Asp Gly Glu Asn Lys
        360                 365                 370

aag gaa gga cga gtg gag gtt ttt gtc aat ggc caa tgg gga aca atc   1253
Lys Glu Gly Arg Val Glu Val Phe Val Asn Gly Gln Trp Gly Thr Ile
    375                 380                 385

tgc gat gac gga tgg acc gat aag cat gca gct gtg atc tgc cgg caa   1301
Cys Asp Asp Gly Trp Thr Asp Lys His Ala Ala Val Ile Cys Arg Gln
390                 395                 400                 405

ctt ggc tat aag ggt cct gcc aga gca agg act atg gct tat ttt ggg   1349
Leu Gly Tyr Lys Gly Pro Ala Arg Ala Arg Thr Met Ala Tyr Phe Gly
                410                 415                 420

gaa gga aaa ggc ccc atc cac atg gat aat gtg aag tgc aca gga aat   1397
Glu Gly Lys Gly Pro Ile His Met Asp Asn Val Lys Cys Thr Gly Asn
            425                 430                 435
```

```
gag aag gcc ctg gct gac tgt gtc aaa caa gac att gga agg cac aac   1445
Glu Lys Ala Leu Ala Asp Cys Val Lys Gln Asp Ile Gly Arg His Asn
        440                 445                 450

tgc cgc cac agt gag gat gca gga gtc atc tgt gac tat tta gag aag   1493
Cys Arg His Ser Glu Asp Ala Gly Val Ile Cys Asp Tyr Leu Glu Lys
    455                 460                 465

aaa gca tca agt agt ggt aat aaa gag atg ctc tca tct gga tgt gga   1541
Lys Ala Ser Ser Ser Gly Asn Lys Glu Met Leu Ser Ser Gly Cys Gly
470                 475                 480                 485

ctg agg tta ctg cac cgt cgg cag aaa cgg atc att ggt ggg aac aat   1589
Leu Arg Leu Leu His Arg Arg Gln Lys Arg Ile Ile Gly Gly Asn Asn
                490                 495                 500

tct tta agg ggt gcc tgg cct tgg cag gct tcc ctc agg ctg agg tcg   1637
Ser Leu Arg Gly Ala Trp Pro Trp Gln Ala Ser Leu Arg Leu Arg Ser
            505                 510                 515

gcc cat gga gac ggc agg ctg ctt tgt gga gct acc ctt ctg agt agc   1685
Ala His Gly Asp Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu Ser Ser
        520                 525                 530

tgc tgg gtc ctg aca gct gca cac tgc ttc aaa agg tac gga aac aac   1733
Cys Trp Val Leu Thr Ala Ala His Cys Phe Lys Arg Tyr Gly Asn Asn
    535                 540                 545

tcg agg agc tat gca gtt cga gtt ggg gat tat cat act ctg gtc cca   1781
Ser Arg Ser Tyr Ala Val Arg Val Gly Asp Tyr His Thr Leu Val Pro
550                 555                 560                 565

gag gag ttt gaa caa gaa ata ggg gtt caa cag att gtg att cac agg   1829
Glu Glu Phe Glu Gln Glu Ile Gly Val Gln Gln Ile Val Ile His Arg
                570                 575                 580

aac tac agg cca gac aga agc gac tat gac att gcc ctg gtt aga ttg   1877
Asn Tyr Arg Pro Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val Arg Leu
            585                 590                 595

caa gga cca ggg gag caa tgt gcc aga cta agc acc cac gtt ttg cca   1925
Gln Gly Pro Gly Glu Gln Cys Ala Arg Leu Ser Thr His Val Leu Pro
        600                 605                 610

gcc tgt tta cct cta tgg aga gag agg cca cag aaa aca gcc tcc aac   1973
Ala Cys Leu Pro Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala Ser Asn
    615                 620                 625

tgt cac ata aca gga tgg gga gac aca ggt cgt gcc tac tca aga act   2021
Cys His Ile Thr Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser Arg Thr
630                 635                 640                 645

cta caa caa gct gct gtg cct ctg tta ccc aag agg ttt tgt aaa gag   2069
Leu Gln Gln Ala Ala Val Pro Leu Leu Pro Lys Arg Phe Cys Lys Glu
                650                 655                 660
```

```
agg tac aag gga cta ttt act ggg aga atg ctc tgt gct ggg aac ctc   2117
Arg Tyr Lys Gly Leu Phe Thr Gly Arg Met Leu Cys Ala Gly Asn Leu
            665                 670                 675

caa gaa gac aac cgt gtg gac agc tgc cag gga gac agt gga gga cca   2165
Gln Glu Asp Asn Arg Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro
        680                 685                 690

ctc atg tgt gaa aag cct gat gag tcc tgg gtt gtg tat ggg gtg act   2213
Leu Met Cys Glu Lys Pro Asp Glu Ser Trp Val Val Tyr Gly Val Thr
    695                 700                 705

tcc tgg ggg tat gga tgt gga gtc aaa gac act cct gga gtt tat acc   2261
Ser Trp Gly Tyr Gly Cys Gly Val Lys Asp Thr Pro Gly Val Tyr Thr
710                 715                 720                 725

aga gtc ccc gct ttt gta cct tgg ata aaa agt gtc acc agt ctg       2306
Arg Val Pro Ala Phe Val Pro Trp Ile Lys Ser Val Thr Ser Leu
                730                 735                 740

taacttatgg aaagctcaag aaatagtaaa acagtaacta ttcagtcttc           2356
```

<210> 4
<211> 761
<212> PRT
<213> Mus musculus

<400> 4

```
Met Ala Leu Ala Arg Cys Val Leu Ala Val Ile Leu Gly Ala Leu Ser
    -20                 -15                 -10

Val Val Ala Arg Ala Asp Pro Val Ser Arg Ser Pro Leu His Arg Pro
 -5              -1   1               5                  10

His Pro Ser Pro Pro Arg Ser Gln His Ala His Tyr Leu Pro Ser Ser
            15                  20                  25

Arg Arg Pro Pro Arg Thr Pro Arg Phe Pro Leu Pro Leu Arg Ile Pro
        30                  35                  40

Ala Ala Gln Arg Pro Gln Val Leu Ser Thr Gly His Thr Pro Pro Thr
    45                  50                  55

Ile Pro Arg Arg Cys Gly Ala Gly Glu Ser Trp Gly Asn Ala Thr Asn
60                  65                  70                  75

Leu Gly Val Pro Cys Leu His Trp Asp Glu Val Pro Pro Phe Leu Glu
                80                  85                  90

Arg Ser Pro Pro Ala Ser Trp Ala Glu Leu Arg Gly Gln Pro His Asn
            95                  100                 105

Phe Cys Arg Ser Pro Asp Gly Ser Gly Arg Pro Trp Cys Phe Tyr Arg
        110                 115                 120
```

```
Asn Ala Gln Gly Lys Val Asp Trp Gly Tyr Cys Asp Cys Gly Gln Gly
    125                 130                 135

Pro Ala Leu Pro Val Ile Arg Leu Val Gly Gly Asn Ser Gly His Glu
140                 145                 150                 155

Gly Arg Val Glu Leu Tyr His Ala Gly Gln Trp Gly Thr Ile Cys Asp
                160                 165                 170

Asp Gln Trp Asp Asn Ala Asp Ala Asp Val Ile Cys Arg Gln Leu Gly
            175                 180                 185

Leu Ser Gly Ile Ala Lys Ala Trp His Gln Ala His Phe Gly Glu Gly
        190                 195                 200

Ser Gly Pro Ile Leu Leu Asp Glu Val Arg Cys Thr Gly Asn Glu Leu
    205                 210                 215

Ser Ile Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn Cys Gly
220                 225                 230                 235

His Lys Glu Asp Ala Gly Val Ser Cys Val Pro Leu Thr Asp Gly Val
                240                 245                 250

Ile Arg Leu Ala Gly Gly Lys Ser Thr His Glu Gly Arg Leu Glu Val
            255                 260                 265

Tyr Tyr Lys Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp Thr Glu
        270                 275                 280

Met Asn Thr Tyr Val Ala Cys Arg Leu Leu Gly Phe Lys Tyr Gly Lys
    285                 290                 295

Gln Ser Ser Val Asn His Phe Asp Gly Ser Asn Arg Pro Ile Trp Leu
300                 305                 310                 315

Asp Asp Val Ser Cys Ser Gly Lys Glu Val Ser Phe Ile Gln Cys Ser
                320                 325                 330

Arg Arg Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp Val Gly
            335                 340                 345

Leu Thr Cys Tyr Pro Asp Ser Asp Gly His Arg Leu Ser Pro Gly Phe
        350                 355                 360

Pro Ile Arg Leu Val Asp Gly Glu Asn Lys Lys Glu Gly Arg Val Glu
    365                 370                 375

Val Phe Val Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly Trp Thr
380                 385                 390                 395

Asp Lys His Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys Gly Pro
                400                 405                 410

Ala Arg Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly Pro Ile
```

```
                415                 420                 425

His Met Asp Asn Val Lys Cys Thr Gly Asn Glu Lys Ala Leu Ala Asp
        430                 435                 440

Cys Val Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser Glu Asp
    445                 450                 455

Ala Gly Val Ile Cys Asp Tyr Leu Glu Lys Lys Ala Ser Ser Ser Gly
460                 465                 470                 475

Asn Lys Glu Met Leu Ser Ser Gly Cys Gly Leu Arg Leu Leu His Arg
                480                 485                 490

Arg Gln Lys Arg Ile Ile Gly Gly Asn Asn Ser Leu Arg Gly Ala Trp
            495                 500                 505

Pro Trp Gln Ala Ser Leu Arg Leu Arg Ser Ala His Gly Asp Gly Arg
        510                 515                 520

Leu Leu Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu Thr Ala
    525                 530                 535

Ala His Cys Phe Lys Arg Tyr Gly Asn Asn Ser Arg Ser Tyr Ala Val
540                 545                 550                 555

Arg Val Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu Gln Glu
                560                 565                 570

Ile Gly Val Gln Gln Ile Val Ile His Arg Asn Tyr Arg Pro Asp Arg
            575                 580                 585

Ser Asp Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Gly Glu Gln
        590                 595                 600

Cys Ala Arg Leu Ser Thr His Val Leu Pro Ala Cys Leu Pro Leu Trp
    605                 610                 615

Arg Glu Arg Pro Gln Lys Thr Ala Ser Asn Cys His Ile Thr Gly Trp
620                 625                 630                 635

Gly Asp Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala Ala Val
                640                 645                 650

Pro Leu Leu Pro Lys Arg Phe Cys Lys Glu Arg Tyr Lys Gly Leu Phe
            655                 660                 665

Thr Gly Arg Met Leu Cys Ala Gly Asn Leu Gln Glu Asp Asn Arg Val
        670                 675                 680

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu Lys Pro
    685                 690                 695

Asp Glu Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr Gly Cys
700                 705                 710                 715
```

```
Gly Val Lys Asp Thr Pro Gly Val Tyr Thr Arg Val Pro Ala Phe Val
                720                 725                 730

Pro Trp Ile Lys Ser Val Thr Ser Leu
            735                 740
```

<210> 5
<211> 257
<212> PRT
<213> Homo sapiens

<400> 5

```
Cys Gly Leu Arg Leu Leu His Arg Arg Gln Lys Arg Ile Ile Gly Gly
  1               5                  10                  15

Lys Asn Ser Leu Arg Gly Gly Trp Pro Trp Gln Val Ser Leu Arg Leu
             20                  25                  30

Lys Ser Ser His Gly Asp Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu
         35                  40                  45

Ser Ser Cys Trp Val Leu Thr Ala Ala His Cys Phe Lys Arg Tyr Gly
     50                  55                  60

Asn Ser Thr Arg Ser Tyr Ala Val Arg Val Gly Asp Tyr His Thr Leu
 65                  70                  75                  80

Val Pro Glu Glu Phe Glu Glu Glu Ile Gly Val Gln Gln Ile Val Ile
                 85                  90                  95

His Arg Glu Tyr Arg Pro Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val
            100                 105                 110

Arg Leu Gln Gly Pro Glu Glu Gln Cys Ala Arg Phe Ser Ser His Val
        115                 120                 125

Leu Pro Ala Cys Leu Pro Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala
    130                 135                 140

Ser Asn Cys Tyr Ile Thr Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser
145                 150                 155                 160

Arg Thr Leu Gln Gln Ala Ala Ile Pro Leu Leu Pro Lys Arg Phe Cys
                165                 170                 175

Glu Glu Arg Tyr Lys Gly Arg Phe Thr Gly Arg Met Leu Cys Ala Gly
            180                 185                 190

Asn Leu His Glu His Lys Arg Val Asp Ser Cys Gln Gly Asp Ser Gly
        195                 200                 205

Gly Pro Leu Met Cys Glu Arg Pro Gly Glu Ser Trp Val Val Tyr Gly
    210                 215                 220

Val Thr Ser Trp Gly Tyr Gly Cys Gly Val Lys Asp Ser Pro Gly Val
```

```
225                 230                 235                 240

Tyr Thr Lys Val Ser Ala Phe Val Pro Trp Ile Lys Ser Val Thr Lys
                245                 250                 255

Leu
```

<210> 6
<211> 257
<212> PRT
<213> Mus musculus

<400> 6

```
Cys Gly Leu Arg Leu Leu His Arg Arg Gln Lys Arg Ile Ile Gly Gly
  1               5                  10                  15

Asn Asn Ser Leu Arg Gly Ala Trp Pro Trp Gln Ala Ser Leu Arg Leu
            20                  25                  30

Arg Ser Ala His Gly Asp Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu
        35                  40                  45

Ser Ser Cys Trp Val Leu Thr Ala Ala His Cys Phe Lys Arg Tyr Gly
    50                  55                  60

Asn Asn Ser Arg Ser Tyr Ala Val Arg Val Gly Asp Tyr His Thr Leu
65                  70                  75                  80

Val Pro Glu Glu Phe Glu Gln Glu Ile Gly Val Gln Gln Ile Val Ile
                85                  90                  95

His Arg Asn Tyr Arg Pro Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val
            100                 105                 110

Arg Leu Gln Gly Pro Gly Glu Gln Cys Ala Arg Leu Ser Thr His Val
        115                 120                 125

Leu Pro Ala Cys Leu Pro Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala
    130                 135                 140

Ser Asn Cys His Ile Thr Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser
145                 150                 155                 160

Arg Thr Leu Gln Gln Ala Ala Val Pro Leu Leu Pro Lys Arg Phe Cys
                165                 170                 175

Lys Glu Arg Tyr Lys Gly Leu Phe Thr Gly Arg Met Leu Cys Ala Gly
            180                 185                 190

Asn Leu Gln Glu Asp Asn Arg Val Asp Ser Cys Gln Gly Asp Ser Gly
        195                 200                 205

Gly Pro Leu Met Cys Glu Lys Pro Asp Glu Ser Trp Val Val Tyr Gly
```

210 215 220

Val Thr Ser Trp Gly Tyr Gly Cys Gly Val Lys Asp Thr Pro Gly Val
225 230 235 240

Tyr Thr Arg Val Pro Ala Phe Val Pro Trp Ile Lys Ser Val Thr Ser
245 250 255

Leu

<210> 7
<211> 23
<212> DNA
<213> Mus musculus

<220>
<221> misc_feature
<222> (6)..(18)
<223> Nucleotides 6, 9, 12, 15, and 18 are n wherein n = i.

<400> 7
tgggtnsynw sngcngcnca ttg 23

<210> 8
<211> 20
<212> DNA
<213> Mus musculus

<220>
<221> misc_feature
<222> (9)..(18)
<223> Nucleotides 9, 15, and 18 are n wherein n = i.

<400> 8
acrbtyccnc trwsnccncc 20

<210> 9
<211> 14
<212> PRT
<213> Mus musculus

<400> 9
Ser Ser Cys Trp Val Leu Ser Ala Ala His Cys Phe Leu Glu
1 5 10

<210> 10
<211> 13
<212> PRT
<213> Mus musculus

<400> 10
His Asp Ala Cys Gln Gly Asp Ser Gly Gly Pro Leu Val
1 5 10

<210> 11
<211> 14
<212> PRT
<213> Mus musculus

<400> 11
Ser Pro Cys Trp Val Ala Ser Ala Ala His Cys Phe Ile Gln
1 5 10

<210> 12
<211> 13
<212> PRT
<213> Mus musculus

<400> 12
Thr Asp Ser Cys Lys Gly Asp Ser Gly Gly Pro Leu Ile
1 5 10

<210> 13
<211> 14
<212> PRT
<213> Mus musculus

<400> 13
Ser Asp Arg Trp Val Leu Thr Ala Ala His Cys Ile Leu Tyr
1 5 10

<210> 14
<211> 13
<212> PRT
<213> Mus musculus

<400> 14
Gly Asp Ala Cys Glu Gly Asp Ser Gly Gly Pro Phe Val
1 5 10

<210> 15
<211> 14
<212> PRT
<213> Mus musculus

<400> 15
Ala Pro Glu Trp Val Leu Thr Ala Ala His Cys Leu Lys Ser
1 5 10

<210> 16
<211> 13
<212> PRT
<213> Mus musculus

<400> 16
Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val
1 5 10

<210> 17
<211> 14
<212> PRT
<213> Mus musculus

<400> 17
Asn Asp Gln Trp Val Val Ser Ala Ala His Cys Tyr Lys Tyr
1 5 10

<210> 18
<211> 13
<212> PRT
<213> Mus musculus

<400> 18
Lys Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Val Val
1 5 10

<210> 19
<211> 14
<212> PRT
<213> Mus musculus

<400> 19
Ser Glu Asp Trp Val Val Thr Ala Ala His Cys Gly Val Lys
1 5 10

<210> 20
<211> 13
<212> PRT
<213> Mus musculus

<400> 20
Val Ser Ser Cys Met Gly Asp Ser Gly Gly Pro Leu Val
1 5 10

<210> 21
<211> 14
<212> PRT
<213> Mus musculus

<400> 21
Ala Asn Asn Trp Val Leu Thr Ala Ala His Cys Leu Ser Asn
1 5 10

<210> 22
<211> 13
<212> PRT
<213> Mus musculus

<400> 22
Thr Ser Ser Cys Asn Gly Asp Ser Gly Gly Pro Leu Asn
1 5 10

<210> 23
<211> 32
<212> DNA
<213> EcoRI and BamHI

<220>
<221> misc_feature
<222> (15)..(27)
<223> Nucleotides 15, 18, 21, 24, and 27 are n wherein n = i.

<220>
<221> misc_feature
<222> (16)
<223> Nucleotide 16 is n wherein n c/g.

<220>
<221> misc_feature
<222> (17)
<223> Nucleotide 17 is n wherein n = t/c.

<220>
<221> misc_feature
<222> (19)
<223> Nucleotide 19 is n wherein n = t/a.

<220>
<221> misc_feature
<222> (20)
<223> Nucleotide 20 is n wherein n = g/c.

<220>
<221> misc_feature
<222> (30)
<223> Nucleotide 30 is n wherein n = t/c.

<400> 23
ggggaattct gggtnnnnnn ngcngcncan tg 32

<210> 24
<211> 29
<212> DNA
<213> EcoRI and BamHI

<220>
<221> misc_feature
<222> (12)..(21)
<223> Nucleotides 12, 15, and 21 are n wherein n = i.

<220>
<221> misc_feature
<222> (16)
<223> Nucleotide 16 is n wherein n = g/c.

<220>
<221> misc_feature
<222> (17)
<223> Nucleotide 17 is n wherein n = a/t.

<220>
<221> misc_feature
<222> (18)
<223> Nucleotide 18 is n wherein n = a/g.

<220>
<221> misc_feature
<222> (24)
<223> Nucleotide 24 is n wherein n = c/t.

<220>
<221> misc_feature
<222> (26)
<223> Nucleotide 26 is n wherein = g/c/t.

<220>
<221> misc_feature
<222> (27)
<223> Nucleotide 27 is n wherein n = g/a.

<400> 24
ggggatccc cnccnnnntc nccntnnca 29

<210> 25
<211> 33
<212> DNA
<213> HindIII and XhoI

<220>
<221> misc_feature
<222> (12)..(27)
<223> Nucleotides 12, 21, 24, and 27 are n wherein n = i.

<220>
<221> misc_feature
<222> (15)
<223> Nucleotide 15 is n wherein n = a/g.

<220>
<221> misc_feature
<222> (25)
<223> Nucleotide 25 is n wherein n = a/g.

<220>
<221> misc_feature
<222> (30)
<223> Nucleotide 30 is n wherein n = c/t.

<220>
<221> misc_feature
<222> (33)
<223> Nucleotide 33 is n wherein n = c/t.

<400> 25
gggaagcttg gncantgggg nacnntntgn gan 33

<210> 26
<211> 33
<212> DNA
<213> HindIII and Xhol

<220>
<221> misc_feature
<222> (15)..(28)
<223> Nucleotides 15 and 28 are n wherein n = i.

<400> 26
gggctcgagc cccancctgt tatgtaanag ttg 33

<210> 27
<211> 17
<212> PRT
<213> Mus musculus

<400> 27
Ser Arg Ser Pro Leu His Arg Pro His Pro Ser Pro Pro Arg Ser Gln
1 5 10 15

Xaa

<210> 28
<211> 13
<212> PRT
<213> Mus musculus

<400> 28
Leu Pro Ser Ser Arg Arg Pro Pro Arg Thr Pro Arg Phe
1 5 10

Patent
Attorney's Docket No. 030708-035

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

| | | |
|---|---|---|
| In re Patent Application of | ) | |
| | ) | |
| Peter SONDEREGGER | ) | Group Art Unit: Unassigned |
| | ) | |
| Application No.: | ) | Examiner: Unassigned |
| | ) | |
| Filed: October 26, 1999 | ) | |
| | ) | |
| For: NEUROTRYPSIN | ) | |

**PRELIMINARY AMENDMENT**

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Prior to examination on the merits, please amend the subject application as follows:

IN THE CLAIMS:

Please cancel claims 1-46 without prejudice or disclaimer.

Please add the following new claims 47-61:

-- 47. Neurotrypsins of the formulas I and II

I: neurotrypsin of the human

II: neurotrypsin of the mouse

48. Neurotrypsin according to claim 47, characterized in that the compounds of the formulas I and II comprise the

separate, coding nucleotide sequences and the coded amino acid sequences of the compounds of the formulas I or II.

49. Use of the coding nucleotide sequences of the compounds of the formulas I or II for the production of recombinant proteins.

50. Use of proteins with the coded amino acid sequences of the compounds of the formulas I or II as targets for the development of pharmaceutical drugs, for example for the inhibition or the enhancement of the catalytic activity of the coded proteins of the formulas I or II.

51. Use of the species-homologous proteins of the compounds of the formulas I or II as targets for the development of pharmaceutical drugs, for example for the inhibition or the enhancement of the catalytic activity of the coded proteins of the formulas I or II.

52. Use of the proteins with the coded amino acid sequences of the compounds of the formulas I or II for the

spatial structure determination, for example the spatial structure determination by means of crystallography or nuclear resonance spectroscopy.

53. Use of the coded amino acid sequences of the compounds of the formulas I or II for the prediction of the protein structure by means of computerized protein structure prediction methods.

54. Use of the spatial structure of the coded amino acid sequences of the compounds of the formulas I or II as targets for the development of pharmaceutical drugs, for example for the inhibition or the enhancement of the catalytic activity of the coded proteins of the compounds of the formulas I or II.

55. Use of the coding nucleotide sequences of the compounds of the formulas I or II in gene therapeutical applications in humans and in animals, as for example as parts of gene therapy vectors as for example as parts of artificial chromosomes.

56. Use the compounds of the formulas I or II for so-called cell engineering applications for the production of gene technologically mutated cells, which produce the coded sequences.

57. Use of the coded amino acid sequences of the compounds of the formulas I or II as antigens for the production of antibodies, as for example antibodies that inhibit or promote the protease function or antibodies that can be used for immunohistochemical studies.

58. Use of the coding nucleotide sequences of the compounds of the formulas I or II for the production of transgenic animals, as for example transgenic mice.

59. Use of the coding nucleotide sequences of the compounds of the formulas I or II for the inactivation or the mutation of the corresponding gene by means of gene targeting techniques, as for example the elimination of the gene in the mouse through homologous recombination.

60. Use of the compounds of the formulas I or II for the diagnostics of disorders in the gene corresponding to the compound of the formula I.

61. Use of the coding nucleotide sequences of the compounds of the formulas I or II as a starting sequence for gene technological modifications aimed at the production of pharmaceutical compositions or gene therapy vectors which exhibit changed properties as compared with the corresponding pharmaceutical compositions or gene therapy vectors containing the coding nucleotide sequence of the compounds of formulas I or II, for example changed proteolytic activity, changed proteolytic specificity, or changed pharmacokinetic characteristics.--

REMARKS

Support for the new claims can be found, at least, in original claims 1-46.

Early and favorable consideration of the subject application is earnestly solicited.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

By: Bruce J. Boggs, Jr.
Bruce J. Boggs, Jr.
Registration No. 32,344
by Richard C. Eberhart
Reg No. 37,027

P.O. Box 1404
Alexandria, Virginia 22313-1404
(703) 836-6620

Date: October 26, 1999

# Neurotrypsin

## Technical Field

The present invention is directed to neurotrypsins and to a pharmaceutical composition which contains these substances or has an influence on these substances.

## Disclosure of Invention

Neurotrypsin is a newly discovered serine protease, which is predominantly expressed in the brain and in the lungs; the expression in the brain takes place nearly exclusively in the neurons.

Neurotrypsin has a previously not yet found domain composition: besides the protease domain, there are found 3 or 4 SRCR (scavenger receptor cysteine-rich) domains and one Kringle domain. It is to be pointed out that the combination of Kringle and SRCR domains have not yet been found in proteins. At the amino terminus of the neurotrypsin protein there is a segment of more than 60 amino acids, which has an extremely high proportion of proline and basic amino acids (arginine and histidine).

The invention is characterized by the characteristics in the independent claims. Preferred embodiments are defined in the dependent claims.

The newly found neurotrypsins

- neurotrypsin of the human (compound of the formula I),
- neurotrypsin of the mouse (compound of the formula II)

differ structurally very much from the so far known serine proteases.

The serine protease whose protease domain is structurally most closely related with the protease domain of the new compounds, namely plasmin (of the human), has only a 44 % amino acid sequence identity.

The proline-rich, basic segment at the amino terminus has a certain resemblance with the basic segments of the netrins and the semaphorins/collapsins. Due to this

segment, it is probable that neurotrypsin may be enriched by means of heparin-affinity chromatography.

The neurotrypsins of the human (compound of the formula I) and of the mouse (compound of the formula II) exhibit a very high structural similarity among each other.

The identity of the amino acid sequences of the native proteins of the compounds of the formulas I or II amounts to 81%.

The neurotrypsin of the human (compound of the formula I) has a coding sequence of 2625 nucleotides. The coded peptide of the compound of the formula I has a length of 875 amino acids and contains a signal peptide of 20 amino acids. The neurotrypsin of the mouse (compound of the formula II) has a coding sequence of 2283 nucleotides. The coded protein of the compound of the formula II has a length of 761 amino acids and contains a signal peptide of 21 amino acids. The reason for the greater length of the neurotrypsin of the human consists therein that the human neurotrypsin has 4 SRCR domains, whereas the neurotrypsin of the mouse has only 3 SRCR domains.

The domains which are present in both compounds (compound of the formula I and compound of the formula II) have a high degree of sequence similarity. The corresponding SRCR domains of the compounds of the formulas I and II have an amino acid sequence identity from 81% to 91%. The corresponding Kringle domains have an amino acid sequence identity of 75%. A high degree of similarity consists also in the enzymatically active (i.e. proteolytic) domain (90% amino acid sequence identity).

The protease domains of the neurotrypsins of the human (compound of the formula I) and of the mouse (compound of the formula II) are aligned in the following section, in order to illustrate the high degree of sequence identity.

```
CGLRLLHRRQKRIIGGKNSLRGGWPWQVSLRLKSSHGDGRLLCGATLLSS      50
CGLRLLHRRQKRIIGGNNSLRGAWPWQASLRLRSAHGDGRLLCGATLLSS

CWVLTAAHCFKRYGNSTRSYAVRVGDYHTLVPEEFEEEIGVQQIVIHREY      100
CWVLTAAHCFKRYGNNSRSYAVRVGDYHTLVPEEFEQEIGVQQIVIHRNY

RPDRSDYDIALVRLQGPEEQCARFSSHVLPACLPLWRERPQKTASNCYIT      150
RPDRSDYDIALVRLQGPGEQCARLSTHVLPACLPLWRERPQKTASNCHIT

GWGDTGRAYSRTLQQAAIPLLPKRFCEERYKGRFTGRMLCAGNLHEHKRV      200
GWGDTGRAYSRTLQQAAVPLLPKRFCKERYKGLFTGRMLCAGNLQEDNRV

DSCQGDSGGPLMCERPGESWVVYGVTSWGYGCGVKDSPGVYTKVSAFVPW      250
DSCQGDSGGPLMCEKPDESWVVYGVTSWGYGCGVKDTPGVYTRVPAFVPW

IKSVTKL                                                  258
IKSVTSL
```

From the 258 amino acid sequence positions included in the comparison there are 233 amino acids that are identical in both compounds (upper sequence: compound of the formula I; lower sequence: compound of the formula II; identical amino acids are indicated by vertical lines).

The inventive neurotrypsins are unique when compared with the known serine proteases in that they are expressed according to currently available observations in a distinct degree in neurons. A further organ with a strong expression of neurotrypsin are the lungs (see Gschwend et al., Mol. Cell. Neurosci. 9, pages 207-219, 1997).

The proteins that are structurally most similar to the compounds of the formulas I or II are serine proteases, such as tissue-type plasminogen activator (tPA), urokinase-type plasminogen activator (uPA), plasmin, trypsin, apolipoprotein (a), coagulation factor XI, neuropsin, and acrosin.

In the adult brain, the inventive compounds are expressed predomiantly in the cerebral cortex, the hippocampus, and the amygdala.

In the adult brain stem and the spinal cord, the inventive compounds are expressed predominantly in the motor neurons. A slightly weaker expression is found in the neurons of the superficial layers of the dorsal horn of the spinal cord.

In the adult peripheral nervous system, the inventive compounds are expressed in a subpopulation of the sensory ganglia neurons.

The inventive compounds were found in connection with a study aimed at discovering trypsin-like serine proteases in the nervous system.

The first compound that was found and characterized was the compound of the formula II (Gschwend et al., Mol. Cell. Neurosci. 9, pages 207-219, 1997).

By means of an alignment of the protease domains of 7 known serine proteases (tissue-type plasminogen activator, urokinase-type plasminogen activator, thrombin, plasmin, trypsin, chymotrypsin, and pancreatic elastase) in the proximity of the histidine and the serine of the catalytic triade of the active site, the sequences of the so-called primer oligonucleotides for the polymerase chain reaction were determined.

The primer oligonucleotides were used in a polymerase chain reaction (PCR) together with ss-cDNA from total RNA of the brains of 10 days old mice and resulted in the amplification of a cDNA fragment of a length of approximately 500 base pairs.

This cDNA fragment was used successfully for the isolation of further cDNA fragments by screening commercially available cDNA libraries. Together, the isolated cDNA fragments covered the full length of the coding part of the compound of the formula II.

By conventional DNA sequencing the complete nucleotide sequence and the amino acid sequence deduced therefrom was obtained.

The compound of the formula I was cloned based on its pronounced similarity with the compound of the formula II.

The primer oligonucleotides used were synthesized according to the known sequence of the compound of the formula II.

The cloning of the compound of the formula I was performed by means of two commercially available cDNA libraries from fetal human brain.

This procedure for the cloning can also be used for the isolation of the homologous compounds of other species, such as rat, rabbit, guinea pig, cow, sheep, pig, primates, birds, zebra fish (Brachydanio rerio), Drosophila melanogaster, Caenorhabditis elegans etc.

The coding nucleotide sequences can be used for the production of proteins with the coded amino acid sequences of the compounds of the formulas I or II. A procedure developed in our laboratory allows the production of recombinant proteins in myeloma cells as fusion proteins with an immunoglobulin domain (constant domain of the kappa light chain). The principle of the construction is given in detail by Rader et al. (Rader et al., Eur. J. Biochem. 215, pages 133-141, 1993). The fusion protein produced by the myeloma cells was isolated by immunoaffinity chromatography using a monoclonal antibody against the Ig domain of the kappa light chain. With the same expression method, also the native protein of a compound, starting from the coding sequence, can be produced.

The coding sequences of the compounds of the formulas I or II can be used as starting compounds for the discovery and the isolation of alleles of the compounds of the formulas I or II. Both the polymerase chain reaction and the nucleic acid hybridization can be used for this purpose.

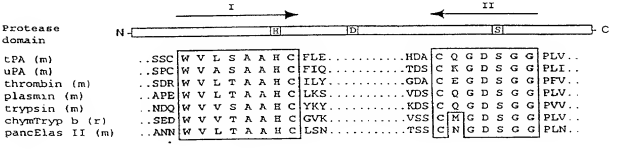
The coding sequences of the compounds of the formulas I or II can be used as starting compounds for so-called "site-directed mutagenesis", in order to generate nucleotide sequences coding the coded proteins that are defined by the compounds of the formulas I or II, or parts thereof, but whose nucleotide sequence is degenerated with respect to the compounds of the formulas I or II due to use of alternative codons.

The coding sequences of the compounds of the formulas I or II can be used as starting compounds for the production of sequence variants by means of so-called site-directed mutagenesis.

Best Modes for Carrying out the Invention (Examples)

cDNA cloning of the compound of the formula II (neurotrypsin of the mouse)

Total RNA was isolated from the brains of 10 days old mice (ICR-ZUR) according to the method of Chomczynski and Sacchi (1987). The production of single stranded cDNA was carried out using oligo(dT) primer and a RNA-dependent DNA polymerase (SuperScript RNase H$^-$-Reverse Transcriptase; Gibco BRL, Gaithersburg, MD) according to the instruction of the supplier. For the realization of the polymerase chain reaction one forward primer was synthesized based on the amino acid sequence of the region of the conserved histidine of the catalytic triade and one primer in the backward direction was synthesized based on the amino acid sequence of the region of the conserved serine of the catalytic triade of the serine proteases. The amino acid sequences used for the determination of the oligonucleotide primers were taken from seven known serine proteases. They are presented in the following.




The protease domains of 7 known serine proteases (tissue-type plasminogen activator, urokinase-type plasminogen activator, thrombin, plasmin, trypsin, chymotrypsin, and pancreatic elastase) were aligned in the region of the conserved histidine and serine of the catalytic triade of the active site. The conserved amino acids of these regions were taken as the basis for the determination of the degenerated primers. The primer sequences are given according to the recommendation of the IUB nomenclature (Nomenclature Committee 1985).

The primers used in the PCR contained restriction sites for *Eco*RI and *Bam*HI at their 5' ends in order to facilitate a subsequent cloning.

The following primers were used:

In the reading direction (sense primers):

5'-GGGGAATTCTGGGTI(C/G)(T/C)I(T/A)(G/C)IGCIGCICA(T/C)TG-3'

In the counter direction (antisense primers):

5'-GGGGGATCCCCICCI(G/C)(A/T)(A/G)TCICC(C/T)T(G/C/T)(G/A)CA-3'.

The polymerase chain reaction was carried out under standard conditions using the DNA polymerase AmpliTaq (Perkin Elmer) according to the recommendations of the producer. The following PCR profile was employed: 93°C for 3 minutes, followed by 35 cycles of 93°C for 1 minute, 48°C for 2 minutes, and 72°C for 2 minutes. Following the last cycle, the incubation was continued at 72°C for further 10 minutes.

The amplified fragments had an approximate length of 500 base pairs. They were cut with *Eco*RI and *Bam*HI and inserted in a Blue Script vector (Bluescript SK(-), Stratagene). The resulting clones were analyzed by DNA sequence determination using the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. USA 77, pages 2163-2167, 1977) on an automated DNA sequencer (LI-COR, model 4000L; Lincoln, NE) using a commercial sequencing kit (SequiTerm long-read cycle sequencing kit-LC; Epicentre Technologies, Madison, WI). The analysis yielded a sequence of 474 base pairs of the catalytic region of the serine protease domain of the compound of the formula II.

The 474 base pair long PCR fragment was used for screening of an oligo(dT)-primed Uni-ZAP-XR cDNA library from the brain of 20 days old mice (Stratagene; cat. no. 937 319). At total of $3 \times 10^6$ lambda plaques were screened under high stringent conditions (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989) using a radioactively labeled PCR fragment as a probe and 24 positive clones were found.

From the positive Lambda-Uni-ZAP-XR phagemid clones the corresponding Bluescript plasmid was cut out by *in vivo* excision according to a standard method recommended by the producer (Stratagene). In order to determine the length of the inserted fragments the corresponding Bluescript plasmid clones were digested with *Sac*I and *Kpn*I. The clones containing the longest fragments were analyzed by DNA

sequencing (as described above) and for subsequent data analysis the GCG software (version 8.1, Unix; Silicon Graphics, Inc.) was used.

Because none of the clones contained the coding sequence in full length, a second cDNA library was screened. The library used in this screen was an oligo(dT)- and random-primed cDNA library in a Lambda phage (Lambda gt10) which was based on mRNA from 15 days old mouse embryos (oligo(dT)- and random-primed Lambda gt10 cDNA library; Clontech, Palo Alto, CA; cat. no. ML 3002a). As a probe a radioactively labeled DNA fragment (AvaI/AatII) from the 5' end of the longest clone of the first screen was used and approximately $2x10^6$ plaques were screened. This screen resulted in 14 positive clones. The cDNA fragments were excised with *Eco*RI and cloned into the Bluecript vector (KS(+); Stratagene). The sequence analysis was carried out as described above.

In this way the nucleotide sequence over the full length cDNA of 2361 and 2376 base pairs, respectively, of the compound of the formula II was obtained. With the described procedure of PCR cloning it is possible to find and isolate also variant forms of the compounds of the formulas I or II, as for example their alleles or their splice variants. The described method of screening of a cDNA library allows also the detection and the isolation of compounds which hybridize under stringent conditions with the coding sequences of the compounds of the formulas I or II.

Cloning of the cDNA of the compound of the formula I (neurotrypsin of the human)

The cloning of the cDNA of the compound of the formula I was carried out basing on the nucleotide sequence of the compound of the formula II. As a first step, a fragment of the compound of the formula I was amplified using the polymerase chain reaction (PCR). As a matrix we used the DNA obtained from a cDNA library from the brain of a human fetus ($17^{th}$ - $18^{th}$ week of pregnancy) which is commercially available (Oligo(dT)- and random-primed, human fetal brain cDNA library in the Lambda ZAP II vector, cat. no. 936206, Stratagene). The synthetic PCR primers contained restriction sites for *Hind*III and *Xho*I at the 5' end in order to facilitate the subsequent cloning.

In the reading direction (sense primers):

5'-GGGAAGCTTGGICA(A/G)TGGGGIACI(A/G)TITG(C/T)GA(C/T)-3'

In the counter direction (antisense primers):

5'-GGGCTCGAGCCCCAICCTGTTATGTAAIAGTTG-3'

The PCR was carried out under standard conditions using the DNA polymerase Amplitaq (Perkin Elmer) according to the recommendations of the producer. The resulting fragment of 1116 base pairs was inserted into the Bluescript vector (Bluescript SK(-), Stratagene). A 600 base pairs long *Hind*III/*Stu*I fragment, corresponding to the 5' half the 1116 base pairs long PCR fragment, was used for the screening of a Lamda cDNA library from human fetal brain (Human Fetal Brain 5'-STRETCH PLUS cDNA library; Lambda gt10; cat. no. HL 3003 a; Clontech). $2x10^6$ Lambda plaques were screened under high stringent conditions (Sambrook et al., Molecular Cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, 1989) by means of a radioactively labeled PCR fragment, and 23 positive clones were found and isolated.

From the positive Lambda gt10 clones the corresponding cDNA fragments were excised with *Eco*RI and inserted into a Bluescript vector (Bluescript KS(+), Stratagene). The sequencing was carried out by means of the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. USA 77, pages 2163-2167, 1977), using a commercial sequencing kit (SequiTherm long-read cycle sequencing kit-LC; Epicentre Technologies, Madison, WI) and Bluescript-specific primers.

In an alternative sequencing strategy, the cDNA fragments of the positive Lambda gt10 clones were PCR amplified using Lambda-specific primers. The sequencing was carried out as described above.

The computerized analysis of the sequences was performed by means of the program package GCG (version 8.1, Unix; Silicon Graphics Inc.).

In this way the nucleotide sequence over the full length of the cDNA of 3350 base pairs was obtained. With the described procedure for PCR cloning it is possible to find and to isolate also variant forms of the compounds of the formulas I or II, as for example their alleles or their splice variants. The described procedure for the screening of a cDNA library allows also the discovery and the isolation of compounds which hybridize under stringent conditions with the coding sequences of the compounds of the formulas I or II.

Visualization of the coded sequences of the compounds of the formulas I or II by means of antibodies

The more than 60 amino acids long proline-rich, basic segment at the amino terminus of the coded sequence of the compounds of the formulas I or II is well suited for the production of antibodies by means of synthesizing peptides and using them for immunization. We have selected two peptide sequences with a length of 19 and 13 amino acids from the proline-rich, basic segment at the amino terminus of the coded sequence of the compound of the formula II for the generation of antibodies. The peptides had the following sequences:

Peptide 1: $H_2N$-SRS PLH RPH PSP PRS QX-$CONH_2$

Peptide 2: $H_2N$-LPS SRR PPR TPR F-COOH

The two peptides were synthesized chemically, coupled to a macromolecular carrier (Keyhole Limpet Hemacyanin), and injected into 2 rabbits for immunization. The resulting antisera exhibit a high antibody titer and could successfully be used both for the identification of native neurotrypsin in brain extract of the mouse and for the identification of recombinant neurotrypsin. The employed procedure for the generation of antibodies can also be used for the generation of antibodies against the coded sequence of the compound of the formula I.

The resulting antibodies against the partial sequences of the coded sequences of the compounds of the formulas I or II can be used for the detection and the isolation of variant forms of the compounds of the formulas I or II, as for example alleles or splice variants. Such antibodies can also be used for the detection and isolation of gene technologically generated variants of the compounds of the formulas I or II.

Purification of the coded sequences of the compounds of the formulas I or II

Besides conventional chromatographic methods, as for example ion exchange chromatography, the purification of the coded sequences of the compounds of the formulas I or II can also be achieved using two affinity chromatographic purification procedures. One affinity chromatographic purification procedure is based on the availability of antibodies. By coupling the antibodies on a chromatographic matrix, a purification procedure results, in which a very high degree of purity of the corresponding compound can be achieved in one step.

Another important feature that can be used for the purification of the coded sequences of the compounds of the formulas I or II is the proline-rich, basic segment at the amino terminus. It may be expected that, due to the high density of positive charges, this segment mediates the binding of the coded sequences of the compounds of the formulas I or II to heparin and heparin-like affinity matrices. This principle allows also the isolation, or at least the enrichment, of variant forms of the coded sequences of the compounds of the formulas I or II, as for example their alleles or splice variants. Likewise the heparin affinity chromatography can be used for the isolation, or at least the enrichment, of species-homologous proteins of the compounds of the formulas I or II.

Industrial Applicability

The coding sequences of the formulas I and II can be used for the production of the coded proteins or parts thereof of the formulas I and II. The production of the coded proteins can be achieved in procaryotic or eucaryotic expression systems.

The gene expression pattern of the inventive compounds in the brain is extremely interesting, because these molecules are expressed in the adult nervous system predominantly in neurons of those regions that are thought to play an important role in learning and memory functions. Together with the recently found evidence for a role of extracellular proteases in neural plasticity, the expression pattern allows the assumption that the proteolytic activity of neurotrypsin has a role in structural reorganizations in connection with learning and memory operations, for example operations which are involved in the processing and storage of learned behaviors, learned emotions, or memory contents. The inventive compounds may, thus, represent a target for pharmaceutical intervention in malfunctions of the brain.

The gene expression pattern of the inventive compounds in the cerebral cortex (especially layers V and VI) is extremely interesting, because a reduction of the cellular differentiation in the cerebral cortex has been found to be associated with schizophrenia. The inventive compounds may, thus, be a target for pharmaceutical intervention in schizophrenia and related psychiatric diseases.

The coding sequences of the inventive compounds have been found to be increased in the neurons located adjacent to the damaged tissue of a focal ischemic stroke, indicating that the inventive compounds play a role in the tissue reaction in the injured cerebral tissue. The inventive compounds may, thus, represent a target for pharmaceutical intervention after ischemic stroke and other forms of neural tissue damage.

Tissue-type plasminogen activator, a serine protease related to the inventive compounds, has recently been found to be involved in excitotoxicity-mediated neuronal cell death. A similar function is conceivable for the inventive compounds and, thus, the inventive compounds represent a possible target for a pharmacological intervention in diseases in which cell death occurs.

The gene expression pattern of the inventive compounds in the spinal cord and in the sensory ganglia is interesting, because these molecules are expressed in the adult nervous system in neurons of those brain regions that are thought to play a role in the processing of pain, as well as in the pathogenesis of pathological pain. The inventive compounds may, thus, be a target for pharmaceutical intervention in pathological pain.

In the following part statements concerning the compounds of the formulas I or II are given:

(1) INFORMATION ABOUT THE COMPOUND OF THE FORMULA I
(Neurotrypsin of the human)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3350 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single strand
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens
(D) DEVELOPMENT STAGE: fetal
(F) TISSUE TYPE: brain

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: human fetal brain 5'-stretch plus cDNA library in the lambda gt10 vector; catalog No. HL 3003a; Clontech, Palo Alto, CA, USA.

(B) CLONE: cDNA Clone No.:
3-1, 3-2, 3-6, 3-7, 3-8, 3-10, 3-11, 3-12

(ix) FEATURE:

(A) NAME/KEY: Signal peptide
(B) LOCATION: 44 .. 103

(ix) FEATURE:

(A) NAME/KEY: mature peptide
(B) LOCATION: 104 .. 2668

(ix) FEATURE:

(A) NAME/KEY: coding sequence
(B) LOCATION: 44 .. 2668

(ix) FEATURE:

(A) NAME/KEY: Proline-rich, basic segment
(B) LOCATION: 104 .. 319

(ix) FEATURE:

(A) NAME/KEY: Kringle domain
(B) LOCATION: 320 .. 538

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 1
(B) LOCATION: 551 .. 856

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 2
(B) LOCATION: 881 .. 1186

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 3
(B) LOCATION: 1202 .. 1504

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 4
(B) LOCATION: 1541 .. 1846

(ix) FEATURE:

(A) NAME/KEY: proteolytic domain
(B) LOCATION: 1898 .. 2668

(ix) FEATURE:

(A) NAME/KEY: histidine of the catalytic triade
(B) LOCATION: 2069 - 2071

(ix) FEATURE:

(A) NAME/KEY: aspartic acid of the catalytic triade
(B) LOCATION: 2219 - 2221

(ix) FEATURE:

(A) NAME/KEY: serine of the catalytic triade
(B) LOCATION: 2516 .. 2518

(ix) FEATURE:

(A) NAME/KEY: polyA signal
(B) LOCATION: 2873 .. 2878

(ix) FEATURE

(A) NAME/KEY: polyA signal
(B) LOCATION: 3034 .. 3039

(ix) FEATURE:

(A) NAME/KEY: polyA signal
(B) LOCATION: 3215 .. 3220

(ix) FEATURE:

(A) NAME/KEY: 3'UTR
(B) LOCATION: 2669 .. 3350

(ix) FEATURE

(A) NAME/KEY: 5'UTR
(B) LOCATION: 1 .. 43

Compound of the formula I (neurotrypsin of the human)

```
CGGAAGCTGG GGAGCATGGA CCAGACCCCG CAGCGCTGGC ACC ATG ACG CTC GCC      55
                                                 Met Thr Leu Ala
                                                 -20

CGC TTC GTG CTA GCC CTG ATG TTA GGG GCG CTC CCC GAA GTG GTC GGC     103
Arg Phe Val Leu Ala Leu Met Leu Gly Ala Leu Pro Glu Val Val Gly
    -15                 -10                 -5                  -1

TTT GAT TCT GTC CTC AAT GAT TCC CTC CAC CAC AGC CAC CGC CAT TCG     151
Phe Asp Ser Val Leu Asn Asp Ser Leu His His Ser His Arg His Ser
  1               5                  10                  15

CCC CCT GCG GGT CCG CAC TAC CCC TAT TAC CTT CCC ACC CAG CAG CGG     199
Pro Pro Ala Gly Pro His Tyr Pro Tyr Tyr Leu Pro Thr Gln Gln Arg
            20                  25                  30

CCC CCG ACG ACG CGT CCG CCG CCG CCT CTC CCG CGC TTC CCG CGC CCC     247
Pro Pro Thr Thr Arg Pro Pro Pro Pro Leu Pro Arg Phe Pro Arg Pro
        35                  40                  45

CCG CGG GCG CTC CCT GCC CAG CGC CCG CAC GCC CTC CAG GCC GGG CAC     295
Pro Arg Ala Leu Pro Ala Gln Arg Pro His Ala Leu Gln Ala Gly His
    50                  55                  60

ACG CCC CGG CCG CAC CCC TGG GGC TGC CCC GCC GGC GAG CCA TGG GTC     343
Thr Pro Arg Pro His Pro Trp Gly Cys Pro Ala Gly Glu Pro Trp Val
65                  70                  75                  80

AGC GTG ACG GAC TTC GGC GCC CCG TGT CTG CGG TGG GCG GAG GTG CCA     391
Ser Val Thr Asp Phe Gly Ala Pro Cys Leu Arg Trp Ala Glu Val Pro
                85                  90                  95

CCC TTC CTG GAG CGG TCG CCC CCA GCG AGC TGG GCT CAG CTG CGA GGA     439
Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Gln Leu Arg Gly
            100                 105                 110

CAG CGC CAC AAC TTT TGT CGG AGC CCC GAC GGC GCG GGC AGA CCC TGG     487
Gln Arg His Asn Phe Cys Arg Ser Pro Asp Gly Ala Gly Arg Pro Trp
        115                 120                 125

TGT TTC TAC GGA GAC GCC CGT GGC AAG GTG GAC TGG GGC TAC TGC GAC     535
Cys Phe Tyr Gly Asp Ala Arg Gly Lys Val Asp Trp Gly Tyr Cys Asp
    130                 135                 140

TGC AGA CAC GGA TCA GTA CGA CTT CGT GGC GGC AAA AAT GAG TTT GAA     583
Cys Arg His Gly Ser Val Arg Leu Arg Gly Gly Lys Asn Glu Phe Glu
145                 150                 155                 160

GGC ACA GTG GAA GTA TAT GCA AGT GGA GTT TGG GGC ACT GTC TGT AGC     631
Gly Thr Val Glu Val Tyr Ala Ser Gly Val Trp Gly Thr Val Cys Ser
                165                 170                 175

AGC CAC TGG GAT GAT TCT GAT GCA TCA GTC ATT TGT CAC CAG CTG CAG     679
Ser His Trp Asp Asp Ser Asp Ala Ser Val Ile Cys His Gln Leu Gln
            180                 185                 190
```

```
CTG GGA GGA AAA GGA ATA GCA AAA CAA ACC CCG TTT TCT GGA CTG GGC      727
Leu Gly Gly Lys Gly Ile Ala Lys Gln Thr Pro Phe Ser Gly Leu Gly
        195                 200                 205

CTT ATT CCC ATT TAT TGG AGC AAT GTC CGT TGC CGA GGA GAT GAA GAA      775
Leu Ile Pro Ile Tyr Trp Ser Asn Val Arg Cys Arg Gly Asp Glu Glu
    210                 215                 220

AAT ATA CTG CTT TGT GAA AAA GAC ATC TGG CAG GGT GGG GTG TGT CCT      823
Asn Ile Leu Leu Cys Glu Lys Asp Ile Trp Gln Gly Gly Val Cys Pro
225                 230                 235                 240

CAG AAG ATG GCA GCT GCT GTC ACG TGT AGC TTT TCC CAT GGC CCA ACG      871
Gln Lys Met Ala Ala Ala Val Thr Cys Ser Phe Ser His Gly Pro Thr
                245                 250                 255

TTC CCC ATC ATT CGC CTT GCT GGA GGC AGC AGT GTG CAT GAA GGC CGG      919
Phe Pro Ile Ile Arg Leu Ala Gly Gly Ser Ser Val His Glu Gly Arg
            260                 265                 270

GTG GAG CTC TAC CAT GCT GGC CAG TGG GGA ACC GTT TGT GAT GAC CAA      967
Val Glu Leu Tyr His Ala Gly Gln Trp Gly Thr Val Cys Asp Asp Gln
        275                 280                 285

TGG GAT GAT GCC GAT GCA GAA GTG ATC TGC AGG CAG CTG GGC CTC AGT      1015
Trp Asp Asp Ala Asp Ala Glu Val Ile Cys Arg Gln Leu Gly Leu Ser
    290                 295                 300

GGC ATT GCC AAA GCA TGG CAT CAG GCA TAT TTT GGG GAA GGG TCT GGC      1063
Gly Ile Ala Lys Ala Trp His Gln Ala Tyr Phe Gly Glu Gly Ser Gly
305                 310                 315                 320

CCA GTT ATG TTG GAT GAA GTA CGC TGC ACT GGG AAT GAG CTT TCA ATT      1111
Pro Val Met Leu Asp Glu Val Arg Cys Thr Gly Asn Glu Leu Ser Ile
                325 .               330                 335

GAG CAG TGT CCA AAG AGC TCC TGG GGA GAG CAT AAC TGT GGC CAT AAA      1159
Glu Gln Cys Pro Lys Ser Ser Trp Gly Glu His Asn Cys Gly His Lys
            340                 345                 350

GAA GAT GCT GGA GTG TCC TGT ACC CCT CTA ACA GAT GGG GTC ATC AGA      1207
Glu Asp Ala Gly Val Ser Cys Thr Pro Leu Thr Asp Gly Val Ile Arg
        355                 360                 365

CTT GCA GGT GGG AAA GGC AGC CAT GAG GGT CGC TTG GAG GTA TAT TAC      1255
Leu Ala Gly Gly Lys Gly Ser His Glu Gly Arg Leu Glu Val Tyr Tyr
    370                 375                 380

AGA GGC CAG TGG GGA ACT GTC TGT GAT GAT GGC TGG ACT GAG CTG AAT      1303
Arg Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp Thr Glu Leu Asn
385                 390                 395                 400

ACA TAC GTG GTT TGT CGA CAG TTG GGA TTT AAA TAT GGT AAA CAA GCA      1351
Thr Tyr Val Val Cys Arg Gln Leu Gly Phe Lys Tyr Gly Lys Gln Ala
                405                 410                 415

TCT GCC AAC CAT TTT GAA GAA AGC ACA GGG CCC ATA TGG TTG GAT GAC      1399
Ser Ala Asn His Phe Glu Glu Ser Thr Gly Pro Ile Trp Leu Asp Asp
            420                 425                 430
```

```
GTC AGC TGC TCA GGA AAG GAA ACC AGA TTT CTT CAG TGT TCC AGG CGA    1447
Val Ser Cys Ser Gly Lys Glu Thr Arg Phe Leu Gln Cys Ser Arg Arg
        435                 440                 445

CAG TGG GGA AGG CAT GAC TGC AGC CAC CGC GAA GAT GTT AGC ATT GCC    1495
Gln Trp Gly Arg His Asp Cys Ser His Arg Glu Asp Val Ser Ile Ala
    450                 455                 460

TGC TAC CCT GGC GGC GAG GGA CAC AGG CTC TCT CTG GGT TTT CCT GTC    1543
Cys Tyr Pro Gly Gly Glu Gly His Arg Leu Ser Leu Gly Phe Pro Val
465                 470                 475                 480

AGA CTG ATG GAT GGA GAA AAT AAG AAA GAA GGA CGA GTG GAG GTT TTT    1591
Arg Leu Met Asp Gly Glu Asn Lys Lys Glu Gly Arg Val Glu Val Phe
                485                 490                 495

ATC AAT GGC CAG TGG GGA ACA ATC TGT GAT GAT GGA TGG ACT GAT AAG    1639
Ile Asn Gly Gln Trp Gly Thr Ile Cys Asp Asp Gly Trp Thr Asp Lys
            500                 505                 510

GAT GCA GCT GTG ATC TGT CGT CAG CTT GGC TAC AAG GGT CCT GCC AGA    1687
Asp Ala Ala Val Ile Cys Arg Gln Leu Gly Tyr Lys Gly Pro Ala Arg
        515                 520                 525

GCA AGA ACC ATG GCT TAC TTT GGA GAA GGA AAA GGA CCC ATC CAT GTG    1735
Ala Arg Thr Met Ala Tyr Phe Gly Glu Gly Lys Gly Pro Ile His Val
    530                 535                 540

GAT AAT GTG AAG TGC ACA GGA AAT GAG AGG TCC TTG GCT GAC TGT ATC    1783
Asp Asn Val Lys Cys Thr Gly Asn Glu Arg Ser Leu Ala Asp Cys Ile
545                 550                 555                 560

AAG CAA GAT ATT GGA AGA CAC AAC TGC CGC CAC AGT GAA GAT GCA GGA    1831
Lys Gln Asp Ile Gly Arg His Asn Cys Arg His Ser Glu Asp Ala Gly
                565                 570                 575

GTT ATT TGT GAT TAT TTT GGC AAG AAG GCC TCA GGT AAC AGT AAT AAA    1879
Val Ile Cys Asp Tyr Phe Gly Lys Lys Ala Ser Gly Asn Ser Asn Lys
            580                 585                 590

GAG TCC CTC TCA TCT GTT TGT GGC TTG AGA TTA CTG CAC CGT CGG CAG    1927
Glu Ser Leu Ser Ser Val Cys Gly Leu Arg Leu Leu His Arg Arg Gln
        595                 600                 605

AAG CGG ATC ATT GGT GGG AAA AAT TCT TTA AGG GGT GGT TGG CCT TGG    1975
Lys Arg Ile Ile Gly Gly Lys Asn Ser Leu Arg Gly Gly Trp Pro Trp
    610                 615                 620

CAG GTT TCC CTC CGG CTG AAG TCA TCC CAT GGA GAT GGC AGG CTC CTC    2023
Gln Val Ser Leu Arg Leu Lys Ser Ser His Gly Asp Gly Arg Leu Leu
625                 630                 635                 640

TGC GGG GCT ACG CTC CTG AGT AGC TGC TGG GTC CTC ACA GCA GCA CAC    2071
Cys Gly Ala Thr Leu Leu Ser Ser Cys Trp Val Leu Thr Ala Ala His
                645                 650                 655

TGT TTC AAG AGG TAT GGC AAC AGC ACT AGG AGC TAT GCT GTT AGG GTT    2119
Cys Phe Lys Arg Tyr Gly Asn Ser Thr Arg Ser Tyr Ala Val Arg Val
            660                 665                 670
```

```
GGA GAT TAT CAT ACT CTG GTA CCA GAG GAG TTT GAG GAA GAA ATT GGA     2167
Gly Asp Tyr His Thr Leu Val Pro Glu Glu Phe Glu Glu Glu Ile Gly
        675                 680                 685

GTT CAA CAG ATT GTG ATT CAT CGG GAG TAT CGA CCC GAC CGC AGT GAT     2215
Val Gln Gln Ile Val Ile His Arg Glu Tyr Arg Pro Asp Arg Ser Asp
    690                 695                 700

TAT GAC ATA GCC CTG GTT AGA TTA CAA GGA CCA GAA GAG CAA TGT GCC     2263
Tyr Asp Ile Ala Leu Val Arg Leu Gln Gly Pro Glu Glu Gln Cys Ala
705                 710                 715                 720

AGA TTC AGC AGC CAT GTT TTG CCA GCC TGT TTA CCA CTC TGG AGA GAG     2311
Arg Phe Ser Ser His Val Leu Pro Ala Cys Leu Pro Leu Trp Arg Glu
                725                 730                 735

AGG CCA CAG AAA ACA GCA TCC AAC TGT TAC ATA ACA GGA TGG GGT GAC     2359
Arg Pro Gln Lys Thr Ala Ser Asn Cys Tyr Ile Thr Gly Trp Gly Asp
            740                 745                 750

ACA GGA CGA GCC TAT TCA AGA ACA CTA CAA CAA GCA GCC ATT CCC TTA     2407
Thr Gly Arg Ala Tyr Ser Arg Thr Leu Gln Gln Ala Ala Ile Pro Leu
        755                 760                 765

CTT CCT AAA AGG TTT TGT GAA GAA CGT TAT AAG GGT CGG TTT ACA GGG     2455
Leu Pro Lys Arg Phe Cys Glu Glu Arg Tyr Lys Gly Arg Phe Thr Gly
    770                 775                 780

AGA ATG CTT TGT GCT GGA AAC CTC CAT GAA CAC AAA CGC GTG GAC AGC     2503
Arg Met Leu Cys Ala Gly Asn Leu His Glu His Lys Arg Val Asp Ser
785                 790                 795                 800

TGC CAG GGA GAC AGC GGA GGA CCA CTC ATG TGT GAA CGG CCC GGA GAG     2551
Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Glu Arg Pro Gly Glu
                805                 810                 815

AGC TGG GTG GTG TAT GGG GTG ACC TCC TGG GGG TAT GGC TGT GGA GTC     2599
Ser Trp Val Val Tyr Gly Val Thr Ser Trp Gly Tyr Gly Cys Gly Val
            820                 825                 830

AAG GAT TCT CCT GGT GTT TAT ACC AAA GTC TCA GCC TTT GTA CCT TGG     2647
Lys Asp Ser Pro Gly Val Tyr Thr Lys Val Ser Ala Phe Val Pro Trp
        835                 840                 845

ATA AAA AGT GTC ACC AAA CTG TAA TTCTTCATGG AAACTTCAAA GCAGCATTT     2700
Ile Lys Ser Val Thr Lys Leu  *
    850                 855

AAACAAATGG AAAACTTTGA ACCCCCACTA TTAGCACTCA GCAGAGATGA CAACAAATGG  2760

CAAGATCTGT TTTTGCTTTG TGTTGTGGTA AAAAATTGTG TACCCCCTGC TGCTTTTGAG  2820

AAATTTGTGA ACATTTTCAG AGGCCTCAGT GTAGTGGAAG TGATAATCCT TAAATGAACA  2880

TTTTCTACCC TAATTTCACT GGAGTGACTT ATTCTAAGCC TCATCTATCC CCTACCTATT  2940
```

TCTCAAAATC ATTCTATGCT GATTTTACAA AAGATCATTT TTACATTTGA ACTGAGAACC 3000

CCTTTTAATT GAATCAGTGG TGTCTGAAAT CATATTAAAT ACCCACATTT GACATAAATG 3060

CGGTACCCTT TACTACACTC ATGAGTGGCA TATTTATGCT TAGGTCTTTT CAAAAGACTT 3120

GACAAGAAAT CTTCATATTC TCTGTAGCCT TTGTCAAGTG AGGAAATCAG TGGTTAAAGA 3180

ATTCCACTAT AAACTTTTAG GCCTGAATAG GAGTAGTAAA GCCTCAAGGA CATCTGCCTG 3240

TCACAATATA TTCTCAAAGT GATCTGATAT TTGGAAACAA GTATCCTTGT TGAGTACCAA 3300

GTGCTACAGA AACCATAAGA TAAAAATACT TTCTACCTAC AGCGTGCCCG 3350

(1) INFORMATION ABOUT THE COMPOUND OF THE FORMULA II (Neurotrypsin of the mouse)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2376 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single strand
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mus musculus
(D) DEVELOPMENT STAGE: postnatal day 10
(F) TISSUE TYPE: brain
(G) CELL TYPE: neurons

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: mouse brain cDNA library in the lambda Uni-ZAP-XR vector, oligo (dT)-primed, from Balb c mice, postnatal day 20, Cat. No.. 937 319; Stratagene, La Jolla, CA, USA

(B) CLONE: cDNA clone no. 16

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: mouse brain cDNA library in the Lambda gt10 vector, oligo(dT)- and random-primed, embryonic day 15, Cat. No. ML 3002a; Clontech, Palo Alto, CA, USA

(B) CLONE: cDNA clone #25

(ix) FEATURE:

(A) NAME/KEY: signal peptide
(B) LOCATION: 24 .. 86

(ix) FEATURE:

(A) NAME/KEY: mature peptide
(B) LOCATION: 87 .. 2306

(ix) FEATURE:

(A) NAME/KEY: coding sequence
(B) LOCATION: 24 .. 2306

(ix) FEATURE:

(A) NAME/KEY: proline-rich, basic segment
(B) LOCATION: 90 .. 275

(ix) FEATURE:

(A) NAME/KEY: Kringle domain
(B) LOCATION: 276 .. 494

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 1
(B) LOCATION: 519 .. 824

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 2
(B) LOCATION: 840 .. 1142

(ix) FEATURE:

(A) NAME/KEY: SRCR domain 3
(B) LOCATION: 1179 .. 1484

(ix) FEATURE:

(A) NAME/KEY: proteolytic domain
(B) LOCATION: 1536 .. 2306

(ix) FEATURE:

(A) NAME/KEY: histidine of the catalytic triade
(B) LOCATION: 1707 .. 1709

(ix) FEATURE:

(A) NAME/KEY: aspartic acid of the catalytic triade
(B) LOCATION: 1857 .. 1859

(ix) FEATURE:

(A) NAME/KEY: serine of the catalytic triade

(B) LOCATION: 2154 .. 2156

(ix) FEATURE:

(A) NAME/KEY:polyA signal
(B) LOCATION: 2324 .. 2329 and 2331 .. 2336

(ix) FEATURE:

(A) NAME/KEY: polyA segment
(B) LOCATION: 2357 .. 2376

(ix) FEATURE:

(A) NAME/KEY: 3'UTR
(B) LOCATION: 2307 .. 2341 or 2307 .. 2356

(ix) FEATURE:

(A) NAME/KEY: 5'UTR
(B) LOCATION: 1 .. 23

Coumpound of the formula II (neurotrypsin of the mouse)

```
GGACCACACT CGGCGCCGCA GCC ATG GCG CTC GCC CGC TGC GTG CTG GCT GTG      53
                          Met Ala Leu Ala Arg Cys Val Leu Ala Val
                             -20                 -15

ATT TTA GGG GCA CTG TCT GTA GTG GCC CGC GCT GAT CCG GTC TCG CGC      101
Ile Leu Gly Ala Leu Ser Val Val Ala Arg Ala Asp Pro Val Ser Arg
    -10                 -5                  1                   5

TCT CCC CTT CAC CGC CCG CAT CCG TCC CCA CCG CGT TCC CAA CAC GCG      149
Ser Pro Leu His Arg Pro His Pro Ser Pro Pro Arg Ser Gln His Ala
                10                  15                  20

CAC TAC CTT CCC AGC TCG CGG CGG CCA CCC AGG ACC CCG CGC TTC CCG      197
His Tyr Leu Pro Ser Ser Arg Arg Pro Pro Arg Thr Pro Arg Phe Pro
            25                  30                  35

CTC CCG CTG CGG ATC CCC GCT GCC CAG CGC CCG CAG GTC CTC AGC ACC      245
Leu Pro Leu Arg Ile Pro Ala Ala Gln Arg Pro Gln Val Leu Ser Thr
        40                  45                  50

GGG CAC ACG CCC CCG ACG ATT CCA CGC CGC TGC GGG GCA GGA GAG TCG      293
Gly His Thr Pro Pro Thr Ile Pro Arg Arg Cys Gly Ala Gly Glu Ser
    55                  60                  65

TGG GGC AAT GCC ACC AAC CTC GGC GTC CCG TGT CTA CAC TGG GAC GAG      341
Trp Gly Asn Ala Thr Asn Leu Gly Val Pro Cys Leu His Trp Asp Glu
 70                  75                  80                  85

GTG CCG CCC TTC CTG GAG CGG TCG CCC CCG GCC AGT TGG GCT GAG CTG      389
Val Pro Pro Phe Leu Glu Arg Ser Pro Pro Ala Ser Trp Ala Glu Leu
                90                  95                  100

CGA GGG CAG CCG CAC AAC TTC TGC CGG AGC CCG GAT GGC TCG GGC AGA      437
Arg Gly Gln Pro His Asn Phe Cys Arg Ser Pro Asp Gly Ser Gly Arg
            105                 110                 115

CCT TGG TGC TTC TAT CGG AAT GCC CAG GGC AAA GTA GAC TGG GGC TAC      485
Pro Trp Cys Phe Tyr Arg Asn Ala Gln Gly Lys Val Asp Trp Gly Tyr
        120                 125                 130

TGC GAT TGT GGT CAA GGC CCG GCG TTG CCC GTC ATT CGC CTT GTT GGT      533
Cys Asp Cys Gly Gln Gly Pro Ala Leu Pro Val Ile Arg Leu Val Gly
    135                 140                 145

GGG AAC AGT GGG CAT GAA GGT CGA GTG GAG CTG TAC CAC GCT GGC CAG      581
Gly Asn Ser Gly His Glu Gly Arg Val Glu Leu Tyr His Ala Gly Gln
150                 155                 160                 165

TGG GGG ACC ATC TGT GAC GAC CAA TGG GAC AAT GCA GAC GCA GAC GTC      629
Trp Gly Thr Ile Cys Asp Asp Gln Trp Asp Asn Ala Asp Ala Asp Val
                170                 175                 180

ATC TGT AGG CAG CTG GGG CTC AGT GGC ATT GCC AAA GCA TGG CAT CAG      677
Ile Cys Arg Gln Leu Gly Leu Ser Gly Ile Ala Lys Ala Trp His Gln
            185                 190                 195
```

```
GCA CAT TTT GGG GAA GGA TCT GGC CCA ATA TTG TTG GAT GAA GTA CGC      725
Ala His Phe Gly Glu Gly Ser Gly Pro Ile Leu Leu Asp Glu Val Arg
        200                 205                 210

TGC ACC GGA AAC GAG CTG TCA ATT GAG CAA TGT CCA AAG AGT TCC TGG      773
Cys Thr Gly Asn Glu Leu Ser Ile Glu Gln Cys Pro Lys Ser Ser Trp
    215                 220                 225

GGC GAA CAT AAC TGT GGC CAT AAA GAA GAT GCT GGA GTG TCT TGT GTT      821
Gly Glu His Asn Cys Gly His Lys Glu Asp Ala Gly Val Ser Cys Val
230                 235                 240                 245

CCT CTA ACA GAT GGT GTC ATC AGA CTG GCA GGA GGA AAA AGT ACC CAT      869
Pro Leu Thr Asp Gly Val Ile Arg Leu Ala Gly Gly Lys Ser Thr His
                250                 255                 260

GAA GGT CGC CTG GAG GTC TAC TAC AAG GGG CAG TGG GGG ACA GTC TGT      917
Glu Gly Arg Leu Glu Val Tyr Tyr Lys Gly Gln Trp Gly Thr Val Cys
            265                 270                 275

GAT GAT GGC TGG ACT GAG ATG AAC ACA TAC GTG GCT TGT CGA CTG CTG      965
Asp Asp Gly Trp Thr Glu Met Asn Thr Tyr Val Ala Cys Arg Leu Leu
        280                 285                 290

GGA TTT AAA TAC GGC AAA CAG TCC TCT GTG AAC CAT TTT GAT GGC AGC     1013
Gly Phe Lys Tyr Gly Lys Gln Ser Ser Val Asn His Phe Asp Gly Ser
    295                 300                 305

AAC AGG CCC ATA TGG CTG GAT GAC GTC AGC TGC TCA GGA AAA GAA GTC     1061
Asn Arg Pro Ile Trp Leu Asp Asp Val Ser Cys Ser Gly Lys Glu Val
310                 315                 320                 325

AGC TTC ATT CAG TGT TCC AGG AGA CAG TGG GGA AGG CAT GAC TGC AGC     1109
Ser Phe Ile Gln Cys Ser Arg Arg Gln Trp Gly Arg His Asp Cys Ser
                330                 335                 340

CAT AGA GAA GAT GTG GGC CTC ACC TGC TAT CCT GAC AGC GAT GGA CAT     1157
His Arg Glu Asp Val Gly Leu Thr Cys Tyr Pro Asp Ser Asp Gly His
            345                 350                 355

AGG CTT TCT CCA GGT TTT CCC ATC AGA CTA GTG GAT GGA GAG AAT AAG     1205
Arg Leu Ser Pro Gly Phe Pro Ile Arg Leu Val Asp Gly Glu Asn Lys
        360                 365                 370

AAG GAA GGA CGA GTG GAG GTT TTT GTC AAT GGC CAA TGG GGA ACA ATC     1253
Lys Glu Gly Arg Val Glu Val Phe Val Asn Gly Gln Trp Gly Thr Ile
    375                 380                 385

TGC GAT GAC GGA TGG ACC GAT AAG CAT GCA GCT GTG ATC TGC CGG CAA     1301
Cys Asp Asp Gly Trp Thr Asp Lys His Ala Ala Val Ile Cys Arg Gln
390                 395                 400                 405

CTT GGC TAT AAG GGT CCT GCC AGA GCA AGG ACT ATG GCT TAT TTT GGG     1349
Leu Gly Tyr Lys Gly Pro Ala Arg Ala Arg Thr Met Ala Tyr Phe Gly
                410                 415                 420

GAA GGA AAA GGC CCC ATC CAC ATG GAT AAT GTG AAG TGC ACA GGA AAT     1397
Glu Gly Lys Gly Pro Ile His Met Asp Asn Val Lys Cys Thr Gly Asn
            425                 430                 435
```

```
GAG AAG GCC CTG GCT GAC TGT GTC AAA CAA GAC ATT GGA AGG CAC AAC      1445
Glu Lys Ala Leu Ala Asp Cys Val Lys Gln Asp Ile Gly Arg His Asn
        440                 445                 450

TGC CGC CAC AGT GAG GAT GCA GGA GTC ATC TGT GAC TAT TTA GAG AAG      1493
Cys Arg His Ser Glu Asp Ala Gly Val Ile Cys Asp Tyr Leu Glu Lys
    455                 460                 465

AAA GCA TCA AGT AGT GGT AAT AAA GAG ATG CTC TCA TCT GGA TGT GGA      1541
Lys Ala Ser Ser Ser Gly Asn Lys Glu Met Leu Ser Ser Gly Cys Gly
470                 475                 480                 485

CTG AGG TTA CTG CAC CGT CGG CAG AAA CGG ATC ATT GGT GGG AAC AAT      1589
Leu Arg Leu Leu His Arg Arg Gln Lys Arg Ile Ile Gly Gly Asn Asn
                490                 495                 500

TCT TTA AGG GGT GCC TGG CCT TGG CAG GCT TCC CTC AGG CTG AGG TCG      1637
Ser Leu Arg Gly Ala Trp Pro Trp Gln Ala Ser Leu Arg Leu Arg Ser
            505                 510                 515

GCC CAT GGA GAC GGC AGG CTG CTT TGT GGA GCT ACC CTT CTG AGT AGC      1685
Ala His Gly Asp Gly Arg Leu Leu Cys Gly Ala Thr Leu Leu Ser Ser
        520                 525                 530

TGC TGG GTC CTG ACA GCT GCA CAC TGC TTC AAA AGG TAC GGA AAC AAC      1733
Cys Trp Val Leu Thr Ala Ala His Cys Phe Lys Arg Tyr Gly Asn Asn
    535                 540                 545

TCG AGG AGC TAT GCA GTT CGA GTT GGG GAT TAT CAT ACT CTG GTC CCA      1781
Ser Arg Ser Tyr Ala Val Arg Val Gly Asp Tyr His Thr Leu Val Pro
550                 555                 560                 565

GAG GAG TTT GAA CAA GAA ATA GGG GTT CAA CAG ATT GTG ATT CAC AGG      1829
Glu Glu Phe Glu Gln Glu Ile Gly Val Gln Gln Ile Val Ile His Arg
                570                 575                 580

AAC TAC AGG CCA GAC AGA AGC GAC TAT GAC ATT GCC CTG GTT AGA TTG      1877
Asn Tyr Arg Pro Asp Arg Ser Asp Tyr Asp Ile Ala Leu Val Arg Leu
            585                 590                 595

CAA GGA CCA GGG GAG CAA TGT GCC AGA CTA AGC ACC CAC GTT TTG CCA      1925
Gln Gly Pro Gly Glu Gln Cys Ala Arg Leu Ser Thr His Val Leu Pro
        600                 605                 610

GCC TGT TTA CCT CTA TGG AGA GAG AGG CCA CAG AAA ACA GCC TCC AAC      1973
Ala Cys Leu Pro Leu Trp Arg Glu Arg Pro Gln Lys Thr Ala Ser Asn
    615                 620                 625

TGT CAC ATA ACA GGA TGG GGA GAC ACA GGT CGT GCC TAC TCA AGA ACT      2021
Cys His Ile Thr Gly Trp Gly Asp Thr Gly Arg Ala Tyr Ser Arg Thr
630                 635                 640                 645

CTA CAA CAA GCT GCT GTG CCT CTG TTA CCC AAG AGG TTT TGT AAA GAG      2069
Leu Gln Gln Ala Ala Val Pro Leu Leu Pro Lys Arg Phe Cys Lys Glu
                650                 655                 660

AGG TAC AAG GGA CTA TTT ACT GGG AGA ATG CTC TGT GCT GGG AAC CTC      2117
Arg Tyr Lys Gly Leu Phe Thr Gly Arg Met Leu Cys Ala Gly Asn Leu
            665                 670                 675
```

```
CAA GAA GAC AAC CGT GTG GAC AGC TGC CAG GGA GAC AGT GGA GGA CCA      2165
Gln Glu Asp Asn Arg Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro
        680                 685                 690

CTC ATG TGT GAA AAG CCT GAT GAG TCC TGG GTT GTG TAT GGG GTG ACT      2213
Leu Met Cys Glu Lys Pro Asp Glu Ser Trp Val Val Tyr Gly Val Thr
    695                 700                 705

TCC TGG GGG TAT GGA TGT GGA GTC AAA GAC ACT CCT GGA GTT TAT ACC      2261
Ser Trp Gly Tyr Gly Cys Gly Val Lys Asp Thr Pro Gly Val Tyr Thr
710                 715                 720                 725

AGA GTC CCC GCT TTT GTA CCT TGG ATA AAA AGT GTC ACC AGT CTG          2306
Arg Val Pro Ala Phe Val Pro Trp Ile Lys Ser Val Thr Ser Leu
                730                 735                 740

TAACTTATGG AAAGCTCAAG AAATAGTAAA ACAGTAACTA TTCAGTCTTC AAAAAAAAAA    2366

AAAAAAAAAA                                                          2376
```

## Patent claims

1. Neurotrypsins of the formulas I and II

   I: neurotrypsin of the human

   II: neurotrypsin of the mouse

2. Neurotrypsin according to claim 1, characterized in that the compounds of the formulas I or II comprise the separate, coding nucleotide sequences and the coded amino acid sequences of the compounds of the formulas I or II.

3. Use of the coding nucleotide sequences of the compounds of the formulas I or II for the production of recombinant proteins.

4. Use of proteins with the coded amino acid sequences of the compounds of the formulas I or II as targets for the development of pharmaceutical drugs, for example for the inhibition or the enhancement of the catalytic activity of the coded proteins of the formulas I or II.

5. Use of the species-homologous proteins of the compounds of the formulas I or II as targets for the development of pharmaceutical drugs, for example for the enhancement or the inhibition of the catalytic activity of the coded proteins of the formulas I or II.

6. Use of the proteins with the coded amino acid sequences of the compounds of the formulas I or II for the spatial structure determination, for example the spatial structure determination by means of crystallography or nuclear resonance spectroscopy.

7. Use of the coded amino acid sequences of the compounds of the formulas I or II for the prediction of the protein structure by means of computerized protein structure prediction methods.

8. Use of the spatial structure of the coded amino acid sequences of the compounds of the formulas I or II as targets for the development of pharmaceutical drugs, for example for the inhibition or the enhancement of the catalytic activity of the coded proteins of the compounds of the formulas I or II.

9. Use of the coding nucleotide sequences of the compounds of the formulas I or II in gene therapeutical applications in humans and in animals, as for example as parts of gene therapy vectors or as for example as parts of artificial chromosomes.

10. Use of the compounds of the formulas I or II for so-called cell engineering applications for the production of gene technologically mutated cells, which produce the coded sequences.

11. Use of the coded amino acid sequences of the compounds of the formulas I or II as antigens for the production of antibodies, as for example antibodies that inhibit or promote the protease function or antibodies that can be used for immunohistochemical studies.

12. Use of the coding nucleotide sequences of the compounds of the formulas I or II for the production of transgenic animals, as for example transgenic mice

13. Use of the coding nucleotide sequences of the compounds of the formulas I or II for the inactivation or the mutation of the corresponding gene by means of gene

targeting techniques, as for example the elimination of the gene in the mouse through homologous recombination

14. Use of the compounds of the formulas I or II for the diagnostics of disorders in the gene corresponding to the compound of the formula I.

15. Use of the coding nucleotide sequences of the compounds of the formulas I or II as a starting sequence for gene technological modifications aimed at the production of pharmaceutical compositions or gene therapy vectors which exhibit changed properties as compared with the corresponding pharmaceutical compositions or gene therapy vectors containing the coding nucleotide sequence of the compounds of formulas I or II, for example changed proteolytic activity, changed proteolytic specificity, or changed pharmacokinetic characteristics.

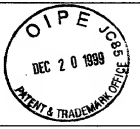
**COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY**
**(Includes Reference to Provisional and PCT International Applications)**

Attorney's Docket No.
030708-035

As a below named inventor, I hereby declare that:
My residence, post office address and citizenship are as stated below next to my name;
I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

NEUROTRYPSIN

the specification of which (check only one item below):




☐ is attached hereto.

☐ was filed as United States application
Number _____
on _____
and was amended
on _____ (if applicable).

☒ was filed as PCT international application
Number PCT/IB98/00625
on April 24, 1998
and was amended
on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 (a)-(e) of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. §119:

| COUNTRY (if PCT, indicate "PCT") | APPLICATION NUMBER | DATE OF FILING (day, month, year) | PRIORITY CLAIMED UNDER 35 U.S.C. §119 |
|---|---|---|---|
| Switzerland | CH966/97 | 26 April 1997 | X Yes __ No |
| | | | __ Yes __ No |
| | | | __ Yes __ No |
| | | | __ Yes __ No |
| | | | __ Yes __ No |

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below.

_____ _____
(Application Number) (Filing Date)

_____ _____
(Application Number) (Filing Date)

**COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (CONTINUED)**
**(Includes Reference to Provisional and PCT International Applications)**

Attorney's Docket No. 030708-035

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose to the Office all information known to me to be material to the patentability as defined in Title 37, Code of Federal Regulations §1.56, which became available between the filing date of the prior application(s) and the national or PCT international filing date of this application:

PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS DESIGNATING THE U.S. FOR BENEFIT UNDER 35 U.S.C. 120:

| U.S. APPLICATIONS | | STATUS *(check one)* | | |
|---|---|---|---|---|
| U.S. APPLICATION NUMBER | U.S. FILING DATE | PATENTED | PENDING | ABANDONED |
| | | | | |
| | | | | |
| | | | | |

| PCT APPLICATIONS DESIGNATING THE U.S. | | | | | |
|---|---|---|---|---|---|
| PCT APPLICATION NO. | PCT FILING DATE | U.S. APPLICATION NUMBERS ASSIGNED (if any) | | | |
| | | | | | |
| | | | | | |
| | | | | | |

I hereby appoint the following attorneys and agent(s) to prosecute said application and to transact all business in the Patent and Trademark Office connected therewith and to file, prosecute and to transact all business in connection with international applications directed to said invention:

| | | | | | |
|---|---|---|---|---|---|
| William L. Mathis | 17,337 | R. Danny Huntington | 27,903 | Gerald F. Swiss | 30,113 |
| Robert S. Swecker | 19,885 | Eric H. Weisblatt | 30,505 | Michael J. Ure | 33,089 |
| Platon N. Mandros | 22,124 | James W. Peterson | 26,057 | Charles F. Wieland III | 33,096 |
| Benton S. Duffett, Jr. | 22,030 | Teresa Stanek Rea | 30,427 | Bruce T. Wieder | 33,815 |
| Norman H. Stepno | 22,716 | Robert E. Krebs | 25,885 | Todd R. Walters | 34,040 |
| Ronald L. Grudziecki | 24,970 | William C. Rowland | 30,888 | Ronni S. Jillions | 31,979 |
| Frederick G. Michaud, Jr. | 26,003 | T. Gene Dillahunty | 25,423 | Harold R. Brown III | 36,341 |
| Alan E. Kopecki | 25,813 | Patrick C. Keane | 32,858 | Allen R. Baum | 36,086 |
| Regis E. Slutter | 26,999 | Bruce J. Boggs, Jr. | 32,344 | Steven M. du Bois | 35,023 |
| Samuel C. Miller, III | 27,360 | William H. Benz | 25,952 | Brian P. O'Shaughnessy | 32,747 |
| Robert G. Mukai | 28,531 | Peter K. Skiff | 31,917 | | |
| George A. Hovanec, Jr. | 28,223 | Richard J. McGrath | 29,195 | | |
| James A. LaBarre | 28,632 | Matthew L. Schneider | 32,814 | | |
| E. Joseph Gess | 28,510 | Michael G. Savage | 32,596 | | |

21839

and: _____

Address all correspondence to:

21839

William L. Mathis
BURNS, DOANE, SWECKER & MATHIS, L.L.P.
P.O. Box 1404
Alexandria, Virginia 22313-1404

Address all telephone calls to: Bruce J. Boggs, Jr. at (703) 836-6620.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

| COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (CONTINUED) (Includes Reference to Provisional and PCT International Applications) | | Attorney's Docket No. 030708-035 |
|---|---|---|
| FULL NAME OF SOLE OR FIRST INVENTOR<br>Peter SONDEREGGER | SIGNATURE | DATE<br>Nov-25-1999 |
| RESIDENCE<br>Zürich, Switzerland CHX | | CITIZENSHIP<br>Swiss |
| POST OFFICE ADDRESS<br>Biochemisches Institut Universität Zürich, Winterthurerstr. 190, CH8057 Zürich, Switzerland | | |
| FULL NAME OF SECOND JOINT INVENTOR, IF ANY | SIGNATURE | DATE |
| RESIDENCE | | CITIZENSHIP |
| POST OFFICE ADDRESS | | |
| FULL NAME OF THIRD JOINT INVENTOR, IF ANY | SIGNATURE | DATE |
| RESIDENCE | | CITIZENSHIP |
| POST OFFICE ADDRESS | | |
| FULL NAME OF FOURTH JOINT INVENTOR, IF ANY | SIGNATURE | DATE |
| RESIDENCE | | CITIZENSHIP |
| POST OFFICE ADDRESS | | |
| FULL NAME OF FIFTH JOINT INVENTOR, IF ANY | SIGNATURE | DATE |
| RESIDENCE | | CITIZENSHIP |
| POST OFFICE ADDRESS | | |
| FULL NAME OF SIXTH JOINT INVENTOR, IF ANY | SIGNATURE | DATE |
| RESIDENCE | | CITIZENSHIP |
| POST OFFICE ADDRESS | | |
| FULL NAME OF SEVENTH JOINT INVENTOR, IF ANY | SIGNATURE | DATE |
| RESIDENCE | | CITIZENSHIP |
| POST OFFICE ADDRESS | | |
| FULL NAME OF EIGHTH JOINT INVENTOR, IF ANY | SIGNATURE | DATE |
| RESIDENCE | | CITIZENSHIP |
| POST OFFICE ADDRESS | | |
| FULL NAME OF NINTH JOINT INVENTOR, IF ANY | SIGNATURE | DATE |
| RESIDENCE | | CITIZENSHIP |
| POST OFFICE ADDRESS | | |